THE ANNALS
of
MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

VOLUME XIX

1948

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves, Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, P. S Dwyer, 116 Rackham Hall, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is \$8.00 inside the Western Hemisphere and \$5.00 elsewhere. Single copies \$3.00. Back numbers are available at \$8.00 per volume or \$3.00 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879

CONTENTS OF VOLUME XIX

Articles

Notes

MISCELLANEOUS

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIX, No. 1 — March, 1948

THE
OF MATHEMA

ON THE GENERALIZED "BIRTH-AND-DEATH" PROCESS

By David G. Kendall

Magdalen College, Oxford

1. Introduction and Summary. The importance of stochastic processes in relation to problems of population growth was pointed out by W. Feller [1] in 1939. He considered among other examples the "birth-and-death" process in which the expected birth and death rates (per head of population per unit of time) were constants, λ_0 and μ_0, say. In this paper I shall give the complete solution of the equations governing the generalised birth-and-death process in which the birth and death rates $\lambda(t)$ and $\mu(t)$ may be any specified functions of the time t. The mathematical method employed starts from M. S. Bartlett's idea of replacing the differential-difference equations for the distribution of the population size by a partial differential equation for its generating function. For an account of this technique,[1] reference may be made to Bartlett's North Carolina lectures [2].

The formulae obtained lead to an expression for the probability of the ultimate extinction of the population, and to the necessary and sufficient condition for a birth-and-death process to be of "transient" type. For transient processes the distribution of the cumulative population is also considered, but here in general it is not found possible to do more than evaluate its mean and variance as functions of t, although a complete solution (including the determination of the asymptotic form of the distribution as t tends to infinity) is obtained for the simple process in which the birth and death rates are independent of the time.

It is shown that a birth-and-death process can be constructed to give an expected population size $\bar{n}_t$ which is any desired function of the time t, and among the many possible solutions the unique one is determined which makes the fluctuation, $\mathrm{Var}(n_t)$, a minimum for all t

The general theory is illustrated with reference to two examples. The first of these is the $(\lambda_0, \mu_1 t)$ process introduced by N. Arley [3] in his study of the cascade showers associated with cosmic radiation; here the birth rate is constant and the death rate is a constant multiple of the "age", t, of the process. The $\bar{n}_t$-curve is then Gaussian in form, and the process is always of transient type.

The second example is provided by the family of "periodic" processes, in which the birth and death rates are periodic functions of the time t. These appear well adapted to describe the response of population growth (or epidemic spread) to the influence of the seasons.

2. The formulation and solution of the equations for the general (λ, μ) process. Let the integer-valued time-dependent random variable n_t measure at time t the

[1] It appears from some remarks by Arley and Borchsenius [5] that the generating function method was first employed in problems of this kind by Dr. C. Palm.

size of a population, and suppose that in an element of time dt the only possible transitions (and their associated probabilities) are:

$$\begin{aligned} n_{t+dt} &= n_t + 1, \qquad && \lambda(t)n_t dt + o(dt); \\ n_{t+dt} &= n_t, && 1 - \{\lambda(t) + \mu(t)\}n_t dt + o(dt); \qquad (1) \\ n_{t+dt} &= n_t - 1, && \mu(t)n_t dt + o(dt). \end{aligned}$$

As an initial condition it will be supposed that the population is descended from a single "ancestor", so that $n_0 = 1$, and thus

$$(2) \qquad P_1(0) = 1, \qquad P_n(0) = 0 \qquad (n \neq 1).$$

It then follows that the $P_n(t)$ must satisfy the differential-difference equations

$$(3) \quad \frac{\partial}{\partial t} P_n(t) = (n+1)\mu P_{n+1}(t) + (n-1)\lambda P_{n-1}(t) - n(\lambda + \mu)P_n(t),\ n \geq 1,$$

and

$$(4) \qquad \frac{\partial}{\partial t} P_0(t) = \mu P_1(t)$$

(where for convenience of writing I have ceased to indicate explicitly the dependence of λ and μ on the time). If $P_n(t)$ is defined to be zero when $n < 0$, the first of the above equations will then be true for all n, and accordingly the generating function

$$(5) \qquad \varphi(z, t) \equiv \sum_{n=-\infty}^{\infty} P_n(t) z^n$$

must satisfy the linear partial differential equation

$$(6) \qquad \frac{\partial \varphi}{\partial t} = (z - 1)(\lambda z - \mu) \frac{\partial \varphi}{\partial z};$$

the problem is to find the solution to this equation when it is coupled with the boundary condition $\varphi(z, 0) = z$.

The equation (6) is of Lagrange's type, and can be solved in the usual manner. The auxiliary equation is

$$(7) \qquad \frac{dz}{dt} = -\mu + (\lambda + \mu)z - \lambda z^2,$$

and while in particular examples it might be convenient to attack this equation directly, progress in general is more easily made by observing that (7) is of Riccati's form, for which a general theory is available.[2] The fundamental property of a Riccati equation is that the general solution is a homographic

[2] See, for example, G. N. Watson [4], pp. 93–94.

function of the constant of integration, so that

$$z = \frac{f_1 + Cf_2}{f_3 + Cf_4},$$

and equally

$$C = \frac{zf_3 - f_1}{f_2 - zf_4},$$

where f_1, f_2, f_3 and f_4 are all functions of the time t. Thus the general solution of (6) is of the form

$$\varphi(z, t) = \Phi\left\{\frac{zf_3 - f_1}{f_2 - zf_4}\right\},$$

and from the boundary condition $\varphi(z, 0) = z$ it then follows that

$$\varphi(z, t) = \frac{g_1(t) + zg_2(t)}{g_3(t) + zg_4(t)}.$$

On expansion, one obtains

(8) $$P_0(t) = \xi_t \text{ and } P_n(t) = \{1 - P_0(t)\}(1 - \eta_t)\eta_t^{n-1} \qquad (n \geqslant 1),$$

where ξ_t and η_t are functions of the time t. Thus, *for the general* (λ, μ) *process, the population size at any time is distributed in a geometric series with a modified zero term*

The next stage of the solution is to determine the functions ξ_t and η_t. From (8),

(9) $$\varphi(z, t) = \frac{\xi + (1 - \xi - \eta)z}{1 - \eta z},$$

and if this expression for φ be substituted in (6) it will be found[3] that

$$(\eta\xi' - \xi\eta') + \eta' = \lambda(1 - \xi)(1 - \eta),$$

and

$$\xi' = \mu(1 - \xi)(1 - \eta).$$

Now let $U = 1 - \xi$ and $V = 1 - \eta$, so that

$$U'/U = -\mu V,$$

and

$$V' = (\mu - \lambda)V - \mu V^2.$$

The last equation is of Bernoulli's type and can be solved by writing

$$W = 1/V,$$

[3] Here $\xi' \equiv d\xi/dt$, etc.

so that

$$W' + (\mu - \lambda)W = \mu.$$

Initially $\xi = \eta = 0$, and $U = V = W = 1$; the solution of the W-equation is therefore

$$W = e^{-\rho}\left\{1 + \int_0^t e^{\rho(\tau)}\mu(\tau)d\tau\right\}, \tag{10a}$$

where the function ρ is defined by

$$\rho(t) = \int_0^t \{\mu(\tau) - \lambda(\tau)\}d\tau. \tag{11}$$

Integration by parts gives two other formulae for W which will prove useful they are

$$W = 1 + e^{-\rho}\int_0^t e^{\rho(\tau)}\lambda(\tau)d\tau, \tag{10b}$$

and

$$W = \tfrac{1}{2}(1 + e^{-\rho}) + \tfrac{1}{2}e^{-\rho}\int_0^t e^{\rho(\tau)}\{\lambda(\tau) + \mu(\tau)\}d\tau. \tag{10c}$$

The quantities U and V, and hence also ξ and η can now be expressed in terms of ρ and W, for

$$\frac{U'}{U} = -\mu V = -\frac{\mu}{W} = -\frac{W'}{W} - \rho',$$

and so

$$\xi_t = 1 - \frac{e^{-\rho}}{W} \quad \text{and} \quad \eta_t = 1 - \frac{1}{W}. \tag{12}$$

These results, together with (8), suffice to determine completely the $P_n(t)$ as functions of the time t.

It is easy to deduce formulae for the mean and variance of n_t (these could also be obtained directly from (6)). For the mean,

$$\bar{n}_t = \frac{1 - \xi_t}{1 - \eta_t} = e^{-\rho(t)}, \tag{13}$$

while for the variance,

$$\begin{aligned} \text{Var}(n_t) &= \frac{(1-\xi)(\xi+\eta)}{(1-\eta)^2} = e^{-\rho}(2W - 1 - e^{-\rho}) \\ &= e^{-2\rho}\int_0^t e^{\rho(\tau)}\{\lambda(\tau) + \mu(\tau)\}d\tau. \end{aligned} \tag{14c}$$

Alternatively, using the other forms for W, one can write

$$\text{(14a)} \qquad \operatorname{Var}(n_t) = e^{-\rho}\left\{e^{-\rho} - 1 + 2e^{-\rho}\int_0^t e^{\rho(\tau)}\mu(\tau)\,d\tau\right\},$$

$$\text{(14b)} \qquad = e^{-\rho}\left\{1 - e^{-\rho} + 2e^{-\rho}\int_0^t e^{\rho(\tau)}\lambda(\tau)\,d\tau\right\}.$$

If the initial population $n_0 = N > 1$, these formulae for $\bar{n}_t$ and $\operatorname{Var}(n_t)$ are to be multiplied by N.

It is now a simple matter to apply these formulae to the Arley $(\lambda_0, \mu_1 t)$ process. It will be found that

$$\rho = \tfrac{1}{2}\mu_1 t^2 - \lambda_0 t.$$

and

$$W = 1 + \lambda_0 e^{-\frac{1}{2}\mu_1 t^2 + \lambda_0 t}\int_0^t e^{\frac{1}{2}\mu_1\tau^2 - \lambda_0\tau}\,d\tau.$$

The mean growth of the process therefore follows the Gaussian law

$$\bar{n}_t = e^{\lambda_0 t - \frac{1}{2}\mu_1 t^2},$$

while for the variance (using (14b), since λ is a constant) one finds

$$\operatorname{Var}(n_t) = \bar{n}_t(1 - \bar{n}_t) + 2\lambda_0\bar{n}_t^2\int_0^t e^{\frac{1}{2}\mu_1\tau^2 - \lambda_0\tau}\,d\tau,$$

in agreement with Arley [3] and Bartlett [2]. The distribution of n_t at time t follows on inserting the above values of ρ and W into (8) and (12).

3. The chances of extinction. The simplest special case is that in which (λ, μ) have the *constant* values (λ_0, μ_0); this is the process introduced by Feller [1] and later discussed by several writers.[4] The formulae (13) and (14c) give at once the results

$$\text{(15)} \qquad \bar{n}_t = e^{(\lambda_0 - \mu_0)t} \quad \text{and} \quad \operatorname{Var}(n_t) = \frac{\lambda_0 + \mu_0}{\lambda_0 - \mu_0}\,\bar{n}_t(\bar{n}_t - 1),$$

due to Feller, while since

$$W = \frac{\lambda_0\bar{n}_t - \mu_0}{\lambda_0 - \mu_0},$$

equations (8) and (12) give

$$\text{(16)} \qquad P_0(t) = \frac{\mu_0(\bar{n}_t - 1)}{\lambda_0\bar{n}_t - \mu_0} \quad \text{and} \quad P_n(t) = \{1 - P_0(t)\}(1 - \eta_t)\eta_t^{n-1} \ (n \geqq 1),$$

[4] See Arley [3], Arley and Borchsenius [5], Bartlett [2] and Kendall [6]. Palm's formulae (16) are stated without proof by Arley and Borchsenius, but it appears from their remarks that he used a generating function method probably identical with that later employed by Bartlett and myself.

where

$$\eta_t = \frac{\lambda_0}{\mu_0} P_0(t) = \frac{\lambda_0(\bar{n}_t - 1)}{\lambda_0 \bar{n}_t - \mu_0}.$$

These formulae were first given by C. Palm.[5] They actually hold only if $\lambda_0 \neq \mu_0$; in the case of equality, $W = 1 + \lambda_0 t$, and then

$$\bar{n}_t = 1, \qquad \operatorname{Var}(n_t) = 2\lambda_0 t,$$

$$(17) \qquad P_0(t) = \frac{\lambda_0 t}{1 + \lambda_0 t} \quad \text{and} \quad P_n(t) = \{1 - P_0(t)\}(1 - \eta_t)\eta_t^{n-1} \qquad (n \geqq 1),$$

where $\eta_t = P_0(t)$.

One particularly interesting point is that

$$P_0(t) \to 1 \text{ as } t \to \infty \text{ if } \lambda_0 \leqq \mu_0,$$

so that the population is "almost certain" to die out, even though in the critical case ($\lambda_0 = \mu_0$) the *expected* population size $\bar{n}_t$ has a constant value. The same is true for any initial size of population; the new expression for $P_0(t)$ is then simply equal to the former one raised to the power $n_0 = N$, and therefore tends to unity as before. This phenomenon of extinction was first noticed in a similar problem[6] by Francis Galton and H. W. Watson; an account of their work is given in Appendix F of Galton's book [7].

The formulae of the last section now make possible a discussion of the chances of extinction for the general (λ, μ) process. When $n_0 = 1$,

$$(18) \qquad P_0(t) = \frac{\int_0^t e^{\rho} \mu \, d\tau}{1 + \int_0^t e^{\rho} \mu \, d\tau},$$

and so *the necessary and sufficient condition for the ultimate extinction of the population is that the integral*

$$(19) \qquad I = \int_0^\infty e^{\rho(\tau)} \mu(\tau) \, d\tau$$

should be divergent.

It will be noticed that the integrand of (19) is non-negative, and so the integral must either diverge to plus infinity, or have a finite value. Hence in any case *the population always has a definite chance of extinction, given by $I/(1 + I)$.* For a population descended from N initial ancestors, the $P_n(t)$ are generated by the function

$$(20) \qquad \left\{ \frac{\xi + (1 - \xi - \eta) z}{1 - \eta z} \right\}^N,$$

[6] The extinction of family-names Further references will be found in my paper [6].

so that

$$P_0(t) = \xi_t^N,$$

and the chance of ultimate extinction is

(21) $$\left(\frac{I}{1+I}\right)^N,$$

which is or is not equal to unity for all N indifferently.

Extinction is impossible, in the sense of being an event of zero probability, if and only if μ is identically zero, so that the process is one of reproduction only. It is also worth noting that a necessary but not sufficient condition for almost certain extinction is the divergence of the integral

(22) $$\int_0^\infty \mu(\tau)d\tau.$$

For if (22) had a finite value, $\rho(t)$ would be bounded for all t, and so (19) could not be divergent. In general, when $I = \infty$ and the population is almost certainly doomed to extinction, I shall speak of the process as *transient*.

For a transient process it is of interest to consider the random variable T, defined to be the "age" of the process at the moment of extinction. Since

$$P_0(t) \equiv \text{Probability } \{T \leq t\},$$

the probability distribution of T is $P_0'(T)dT$, or

(23) $$\frac{e^{\rho(T)}\,\mu(T)dT}{\left\{1 + \int_0^T e^{\rho(\tau)}\,\mu(\tau)d\tau\right\}^2}, \qquad 0 < T < \infty.$$

For example, in the simplest birth-and-death process, when λ and μ are equal constants, the distribution of T is

(24) $$\frac{\lambda_0\, dT}{(1 + \lambda_0 T)^2}, \qquad 0 < T < \infty.$$

This is for an initial population $n_0 = 1$; more generally, when $n_0 = N > 1$, the distribution of T is

$$NP_0'(T)\{P_0(T)\}^{N-1}dT.$$

The *median life-time* T_m is determined by the relation

(25) $$\int_0^{T_m} e^{\rho(\tau)}\mu(\tau)d\tau = 1.$$

For the simple process, $T_m = 1/\lambda_0$ when $\lambda_0 = \mu_0$, and more generally

(26) $$T_m = \frac{1}{\mu_0 - \lambda_0}\cdot \log\left(2 - \frac{\lambda_0}{\mu_0}\right) \qquad (\lambda_0 \neq \mu_0)$$

if $n_0 = 1$. When $n_0 = N > 1$, the formula for T_m becomes

$$\int_0^{T_m} e^{\rho(\tau)} \mu(\tau) d\tau = 1/(2^{1/N} - 1) \sim \frac{N}{\log 2}. \tag{27}$$

For the balanced process (λ_0, λ_0) it therefore follows that

$$T_m(N) = T_m(1)/(2^{1/N} - 1) \sim 1.44\, N\, T_m(1), \tag{28}$$

as N tends to infinity. If the process is unbalanced, however, so that $\lambda_0 < \mu_0$, this asymptotic proportionality to N does not hold, and instead

$$T_m = \frac{1}{\mu_0 - \lambda_0} \log \left\{ \frac{2^{1/N} \mu_0 - \lambda_0}{(2^{1/N} - 1)\mu_0} \right\} \sim \frac{\log N}{\mu_0 - \lambda_0}, \tag{29}$$

as N tends to infinity.

4. The cumulative population. There is associated with a birth-and-death process another random variable, M_t, which is of importance in some applications. This is defined as follows: initially $M_0 = n_0$, while for $t > 0$, M_t shares all the *positive* jumps of n_t.

For example, if n_t represents the number of cases of a disease in a population at time t, M_t will be the total number of cases which have been recorded up to that time. If the process is transient, so that the epidemic is almost certainly extinguished in the course of time, M_∞ will then be a measure of its overall severity.

Again, if n_t represents the *viable count* of a population of bacteria[6] with a birth rate $\lambda(t)$ and a death rate $\mu(t)$, M_t will be equal to the *total count* in which living and dead organisms are not distinguished.

In order to discuss the joint variation of n_t and M_t it is necessary to introduce the new generating function

$$\psi(z, w, t) = \sum_{n=0}^{\infty} \sum_{M=0}^{\infty} P_{n,M}(t) z^n w^M. \tag{30}$$

Here the $P_{n,M}(t)$ give the joint frequency-distribution of n_t and M_t at time t. By the usual argument the differential equation satisfied by the function ψ will be found to be

$$\frac{\partial \psi}{\partial t} = \{\lambda w z^2 - (\lambda + \mu) z + \mu\} \frac{\partial \psi}{\partial z}, \tag{31}$$

and the associated boundary condition (if initially $n_0 = M_0 = 1$) is

$$\psi(z, w, 0) = zw. \tag{32}$$

I have been unable to solve this equation for general $\lambda(t)$ and $\mu(t)$; the solution when λ and μ are constants will be given in the next section. It is however

[6] For some general remarks about birth-and-death processes in relation to bacterial growth, reference may be made to my paper [6].

possible to find general expressions for the mean and variance of M_t; for this purpose it is more convenient[7] to work with the cumulant-generating function

$$K(u, v, t) = \log \psi(e^u, e^v, t). \tag{33}$$

This satisfies the differential equation

$$\frac{\partial K}{\partial t} = \{\lambda(e^{u+v} - 1) - \mu(1 - e^{-u})\} \frac{\partial K}{\partial u}, \tag{34}$$

and of course

$$K = u\bar{n}_t + v\bar{M}_t + \tfrac{1}{2}u^2 \operatorname{Var}(n_t) + \tfrac{1}{2}v^2 \operatorname{Var}(M_t) + uv \operatorname{Cov}(n_t, M_t) + \cdots . \tag{35}$$

Expanding both sides of the equation in powers of u and v, and equating coefficients, one obtains the differential equations

$$\frac{d}{dt}\bar{n}_t = (\lambda - \mu)\bar{n}_t, \tag{36}$$

$$\frac{d}{dt}\operatorname{Var}(n_t) = (\lambda + \mu)\bar{n}_t + 2(\lambda - \mu)\operatorname{Var}(n_t), \tag{37}$$

$$\frac{d}{dt}\bar{M}_t = \lambda\bar{n}_t, \tag{38}$$

$$\frac{d}{dt}\operatorname{Var}(M_t) = \lambda\bar{n}_t + 2\lambda \operatorname{Cov}(n_t, M_t), \tag{39}$$

and

$$\frac{d}{dt}\operatorname{Cov}(n_t, M_t) = \lambda\bar{n}_t + \lambda \operatorname{Var}(n_t) + (\lambda - \mu)\operatorname{Cov}(n_t, M_t). \tag{40}$$

The solutions to the first two equations have of course already been given in section 2; from the third it follows that the mean value of M_t is

$$\bar{M}_t = 1 + \int_0^t e^{-\rho(\tau)}\lambda(\tau)\,d\tau. \tag{41}$$

The solution of the fifth equation is

$$\operatorname{Cov}(n_t, M_t) = \bar{n}_t \int_0^t \left\{1 + \frac{\operatorname{Var}(n_\tau)}{\bar{n}_\tau}\right\} \lambda(\tau)\,d\tau, \tag{42}$$

and so the variance of M_t is

$$\operatorname{Var}(M_t) = \int_0^t \{\bar{n}_\tau + 2\operatorname{Cov}(n_\tau, M_\tau)\}\lambda(\tau)\,d\tau. \tag{43}$$

---

[7] Compare Bartlett [2].

In illustration of these formulae, consider first the Arley $(\lambda_0, \mu_1 t)$ process; from (41)

$$\bar{M}_t = 1 + \lambda_0 \int_0^t e^{\lambda_0 \tau - \frac{1}{2}\mu_1 \tau^2}\, d\tau, \tag{44}$$

but the complete expression for Var (M_t) will be a multiple integral which does not appear to admit of much simplification.

For the simple (λ_0, μ_0) process, however, when $\lambda_0 < \mu_0$, it readily follows that

$$\bar{M}_t = \frac{\mu_0 - \lambda_0 \bar{n}_t}{\mu_0 - \lambda_0}, \tag{45}$$

$$\text{Cov}\,(n_t, M_t) = \frac{\lambda_0 \bar{n}_t}{\mu_0 - \lambda_0}\left\{2\mu_0 t - \frac{\mu_0 + \lambda_0}{\mu_0 - \lambda_0}(1 - \bar{n}_t)\right\}, \tag{46}$$

and

$$\text{Var}\,(M_t) = \frac{\lambda_0(\mu_0 + \lambda_0)}{(\mu_0 - \lambda_0)^2}(1 - \bar{n}_t) - \frac{4\lambda_0^2 \mu_0 t \bar{n}_t}{(\mu_0 - \lambda_0)^2} + \frac{\lambda_0^2(\mu_0 + \lambda_0)}{(\mu_0 - \lambda_0)^3}(1 - \bar{n}_t^2). \tag{47}$$

Thus in the limit, as $t \to \infty$, the mean and variance of M_∞ are

$$\bar{M}_\infty = \frac{\mu_0}{\mu_0 - \lambda_0}, \quad \text{and} \quad \text{Var}\,(M_\infty) = \frac{\lambda_0 \mu_0(\lambda_0 + \mu_0)}{(\mu_0 - \lambda_0)^3}, \tag{48}$$

the covariance of course tending to zero. If the process is balanced, so that $\lambda_0 = \mu_0$ and $\bar{n}_t = 1$, the integral for $\bar{M}_t$ has the value $1 + \lambda_0 t$, which increases without limit as t tends to infinity. This will always be so for a balanced process if the integral

$$\int_0^\infty \lambda(\tau)\, d\tau$$

is divergent.

If the initial population n_0 is equal to $N > 1$, and if all its members are counted into M_0, the only modification necessary to the above formulae is that in each case the right-hand side is to be multiplied by N.

5. The asymptotic distribution of the cumulative population for a simple transient birth-and-death process. The equation (31), which appears in the general case to be intractable even if one only requires the asymptotic distribution determined by $\psi(1, w, \infty)$, can be solved completely in the specially simple case when the birth and death rates $\lambda(t)$ and $\mu(t)$ have the constant values λ_0 and μ_0.

Let α and β be the roots of the quadratic

$$\lambda_0 w z^2 - (\lambda_0 + \mu_0) z + \mu_0 = 0, \tag{49}$$

so chosen that $0 < \alpha < 1 < \beta$; then the general solution of (31) will be found by the usual method to be

$$\psi = \Psi\left\{\frac{z-\alpha}{\beta-z}e^{-\lambda_0 w(\beta-\alpha)t}\right\}.$$

The boundary condition $\psi(z, w, 0) = zw$ therefore gives

$$\psi = w\left(\frac{\alpha(\beta-z)+\beta(z-\alpha)e^{-\lambda_0 w(\beta-\alpha)t}}{(\beta-z)+(z-\alpha)e^{-\lambda_0 w(\beta-\alpha)t}}\right), \tag{50}$$

and it may be noted that if $n_0 = M_0 = N > 1$, this formula for ψ would have to be raised to the Nth power. It will suffice, however, to discuss the simplest case when $n_0 = M_0 = 1$.

Let the process be transient, so that $\lambda_0 \leq \mu_0$; then the asymptotic frequency distribution of M_t when $t \to \infty$ is determined by the generating function

$$\psi(1, w, \infty) = w\alpha = \frac{\lambda_0 + \mu_0 - \sqrt{\{(\lambda_0+\mu_0)^2 - 4\lambda_0\mu_0 w\}}}{2\lambda_0}, \tag{51}$$

and here it is the positive square root which must be taken. The probability distribution of M_∞ is thus

$$Q_M = \frac{\lambda_0+\mu_0}{2\lambda_0}\frac{(2M)!}{2^{2M}(M!)^2}\frac{x^M}{2M-1}, \qquad (M = 1, 2, 3, \cdots), \tag{52}$$

where

$$x = \frac{4\lambda_0\mu_0}{(\lambda_0+\mu_0)^2}. \tag{53}$$

The first few terms are

$$\frac{\mu_0}{\lambda_0+\mu_0}\{1, \tfrac{1}{4}x, \tfrac{1}{8}x^2, \tfrac{5}{64}x^3, \cdots\}, \tag{54}$$

and it is easy to verify that the mean and variance of this distribution agree with the values given in the last section. When $\lambda_0 = \mu_0$, $x = 1$, and then the terms in (54) fall off to zero like $M^{-3/2}$, $\bar{M}_\infty$ being infinite (in accordance with the remarks at the end of section 4).

6. The determination of the process when its mean growth, $\bar{n}_t$, is given. Since $\bar{n}_t = e^{-\rho(t)}$, it follows that

$$\lambda(t) - \mu(t) = \frac{d}{dt}\log\bar{n}_t, \tag{55}$$

and thus if $\bar{n}_t$ is required to be a given function of the time, the birth and death rates must be chosen in accordance with (55); the only other condition is that for all t, $\lambda(t) \geq 0$ and $\mu(t) \geq 0$.

Arley has pointed out that the simple process ($\lambda(t) = c$, $\mu(t) = 0$) gives a smaller *fluctuation*, Var (n_t), than any other simple process with the same mean

growth, say (λ_0, μ_0) where $\lambda_0 - \mu_0 = c$. This suggests that one should consider the more general question: *if $\bar{n}_t$ is given for all* t, *for which choice of the functions* λ(t) *and* μ(t) *will the fluctuation* Var (n_t) *be a minimum?*

Suppose then that the whole region $t > 0$ consists of three sets of intervals, E_1, E_2 and E_3, and that within an interval of the set E_j,

$\bar{n}_t$ is a decreasing function if $j = 1$,

$\bar{n}_t$ is an increasing function if $j = 2$,

and $\bar{n}_t$ is a constant if $j = 3$.

Then one can write

$$\begin{aligned}\text{Var}\,(n_t) = e^{-2\rho}\Sigma_1[e^{\rho(\tau)}] &+ 2e^{-2\rho}\int_{E_1} e^{\rho(\tau)}\lambda(\tau)d\tau \\ &+ e^{-2\rho}\Sigma_2[-e^{\rho(\tau)}] + 2e^{-2\rho}\int_{E_2} e^{\rho(\tau)}\mu(\tau)d\tau \\ &+ e^{-2\rho}\int_{E_3} e^{\rho(\tau)}\{\lambda(\tau) + \mu(\tau)\}d\tau.\end{aligned}$$

Here the terms involving λ and μ explicitly are all non-negative, and so Var (n_t) will be a minimum for the (unique) choice of λ and μ which makes them all vanish, namely:

$$\begin{aligned}&\text{in } E_1,\ \lambda(t) = 0 \quad \text{and } \mu(t) = -\bar{n}_t'/\bar{n}_t;\\ (56)\qquad &\text{in } E_2,\ \lambda(t) = \bar{n}_t'/\bar{n}_t \text{ and } \mu(t) = 0;\\ &\text{in } E_3,\ \lambda(t) = \mu(t) = 0.\end{aligned}$$

However, when one is looking for a (λ, μ) process with a given $\bar{n}_t$ function, this minimum-fluctuation solution would frequently be an artificial one. For example, suppose it is required that $\bar{n}_t$ shall be a Gaussian curve, reducing to unity when $t = 0$; then

$$\bar{n}_t = e^{\lambda_0 t - \frac{1}{2}\mu_1 t^2}, \tag{57}$$

say, and $\lambda(t) - \mu(t) = \lambda_0 - \mu_1 t$; the most natural solution is then the Arley process,

$$\lambda(t) = \lambda_0, \qquad \mu(t) = \mu_1 t.$$

It is of interest that a (λ, μ) process can be found for which the expected growth follows a logistic law,

$$\bar{n}_t = \frac{\alpha}{1 + (\alpha - 1)e^{-\beta t}} \qquad (\alpha > 1, \beta > 0). \tag{58}$$

According to (55) one must have

$$\lambda(t) - \mu(t) = \frac{(\alpha - 1)\beta}{e^{\beta t} + (\alpha - 1)}.$$

The minimum-fluctuation solution is thus the purely reproductive process

$$\lambda(t) = \frac{(\alpha - 1)\beta}{e^{\beta t} + (\alpha - 1)}, \qquad \mu(t) = 0, \tag{59}$$

which satisfies the relation

$$\lambda(t) = \beta\left(1 - \frac{\bar{n}_t}{\alpha}\right), \tag{60}$$

as might have been expected, since the Verhulst-Pearl-Reed differential equation (which forms the *deterministic* basis for the logistic law) is

$$\frac{1}{n}\frac{dn}{dt} = \beta\left(1 - \frac{n}{\alpha}\right). \tag{61}$$

7. "Periodic" birth-and-death processes. As a further example of the general theory it is worth considering the "periodic" processes for which the expected growth $\bar{n}_t$ is a function of the time which repeats itself with the period $\bar{\omega}$. It will then follow that $\rho(t)$ and so also $\lambda(t) - \mu(t)$ have the period $\bar{\omega}$, while $\rho(t)$ must be zero whenever t is an integer multiple of $\bar{\omega}$. The only cases of interest are those in which λ and μ are separately periodic, and then it can be seen from (14c) that

$$\bar{n} = n_0 \text{ and } \operatorname{Var}(n) = kn_0 \int_0^{\bar{\omega}} e^{\rho(\tau)}\{\lambda(\tau) + \mu(\tau)\}d\tau, \tag{62}$$

whenever $t = k\bar{\omega}$, for every positive integer k. Thus, although the *expected* value of n_t repeats itself regularly, in practice this "periodicity" would be obscured by the rapid increase, with increasing t, in the magnitude of the random fluctuations (as measured by Var (n_t)). Moreover, since

$$\int_0^{k\bar{\omega}} e^{\rho(\tau)}\mu(\tau)d\tau = k\int_0^{\bar{\omega}} e^{\rho(\tau)}\mu(\tau)d\tau,$$

it is clear that the process is necessarily transient, there being unit probability that n_t will ultimately be reduced to zero.

Periodic birth-and-death processes are likely to be of importance in biology; it should be pointed out, however, that this type of process describes the stochastic modification of a *regular* periodicity imposed on the model *from outside*, and it is not to be confused with other stochastic models which themselves generate irregular (non-phase-keeping) oscillations. The models discussed in this section are in fact suitable for the quantitative description of seasonal influences.

Before going into further detail it is natural to specialise the model by assuming that the functions λ and μ are at most *simply* harmonic. If $n_0 = 1$, and since there is to be no damping, one will then have

$$\bar{n}_t = e^{\alpha[\sin\nu(t+c)-\sin\nu c]} \qquad (\alpha > 0), \tag{63}$$

where $\nu\bar{\omega} = 2\pi$, and α and ϵ are amplitude and phase constants, respectively. The functions λ and μ are now to be determined from the relation

$$\lambda - \mu = \alpha\nu \cos \nu(t + \epsilon),$$

and this can be done in many ways. The minimum-fluctuation solution would here be artificial, and it is more natural to select two other solutions,

$$\lambda = \alpha\nu\{1 + \cos \nu(t + \epsilon)\}, \qquad \mu = \alpha\nu, \tag{64}$$

and

$$\lambda = \alpha\nu, \qquad \mu = \alpha\nu\{1 - \cos \nu(t + \epsilon)\}, \tag{65}$$

for further consideration. In the first of these the death rate is constant and the birth rate executes simple-harmonic oscillations, while in the second it is the birth rate which is constant, and the death rate which oscillates. It can be seen that, of all solutions of these two types, (64) and (65) are those with the least value for Var (n_t). From formulae (14a) and (14b) it will be found that, for either process,

$$\text{Var}\ (n) = 4\pi k\alpha I_0(\alpha)e^{\alpha \sin \nu\epsilon} \text{ when } t = k\bar{\omega} \tag{66}$$

where $I_0(\alpha)$ is the Bessel function of zero order, of the first kind and of imaginary argument. (It will be noticed that, whenever t is an integer multiple of $\bar{\omega}$, the distribution of the population size n_t is the same for the two models.) For small oscillations, when $t = k\bar{\omega}$,

$$\text{Var}\ (n) \sim 4\pi k\alpha \text{ as } \alpha \to 0 \tag{67}$$

since $I_0(0) = 1$, while for large oscillations

$$\text{Var}\ (n) \sim 2k(2\pi\alpha)^{\frac{1}{2}}/\bar{n}_{\min} \text{ as } \alpha \to \infty. \tag{68}$$

(Here $\bar{n}_{\min}$ is the minimum value of $\bar{n}_t$.)

The calculation of $P_0(\bar{\omega})$ presents some points of interest. For either model it proves to be

$$\frac{2\pi\alpha I_0(\alpha)e^{\alpha \sin \nu\epsilon}}{1 + 2\pi\alpha I_0(\alpha)e^{\alpha \sin \nu\epsilon}}; \tag{69}$$

this is the probability that a population element, known to be descended from a single individual at time $t = 0$, will have become extinct one year later (if one identifies the oscillations with a seasonal effect). It will be seen that $P_0(\bar{\omega})$ will be least when $\sin \nu\epsilon = -1$, and greatest when $\sin \nu\epsilon = +1$; i.e. when $\bar{n}_t$ is expected to have a minimum, or a maximum, at $t = 0$, respectively. Accordingly it follows that the progeny of a new member of the population is most likely to survive till the following year if the "ancestor" commences its "membership" at a time of year when the population would normally have its minimum value.

In conclusion, I wish to thank Professor M. S. Bartlett for many helpful discussions on the subject of this paper.

REFERENCES

[1] W. Feller, "Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung", *Acta Biotheoretica*, Vol. 5 (1939), pp. 11-40.

[2] M. S. Bartlett, *Stochastic Processes* (notes of a course given at the University of North Carolina in the Fall Quarter, 1946). It is understood that copies of these notes are available on request.

[3] N. Arley, *On the Theory of Stochastic Processes and their Application to the Theory of Cosmic Radiation*, G. E. C. Gads Forlag, Copenhagen, 1943, pp. 106-114.

[4] G. N. Watson, *The Theory of Bessel Functions*, University Press, Cambridge, England, 1944.

[5] N. Arley and V. Borchsenius, "On the theory of infinite systems of differential equations and their application to the theory of stochastic processes and the perturbation theory of quantum mechanics", *Acta Mathematica*, Vol. 76 (1945), pp. 261-322 (esp. 298-9).

[6] D. G. Kendall, "On some modes of population growth leading to R. A. Fisher's logarithmic series distribution". To appear in *Biometrika*.

[7] Francis Galton, *Natural Inheritance*, Macmillan, London, 1889.

PROBABILITY OF COINCIDENCE FOR TWO PERIODICALLY RECURRING EVENTS[1]

By Paul I. Richards

Brookhaven National Laboratory

Summary. This paper contains a study of the following problem: Each of two events recurs with definitely known period and duration, while the starting time of each event is unknown. It is desired that, before the elapse of a certain time, the events occur simultaneously and that this "overlap" be of at least a given minimum duration.

The probability of this satisfactory coincidence is first evaluated, and it is found that the solution, while mathematically adequate, is of no value for practical application. This circumstance arises from the possibility that, with certain rational ratios of the periods, the events may "lock in step". Accordingly, an attempt is made to smooth the probability function with respect to small variations in the ratio of the periods. Due to difficulties in manipulating the number-theoretic expressions involved, this smoothing is carried through only by the use of certain approximations. Moreover, because of these same difficulties, an averaged value of the probability itself is not obtained, but, in its stead, there is derived a formula for that fraction of randomly related repeated trials in which the original probability will be less than one-half.

Thus, the original problem is not completely solved. The results obtained, however, do allow one to compare the relative advantages of different situations and to make a rough estimate of the likelihood of success. Generally speaking, the analysis is applicable whenever the ratio of "on time" to "off time" is small for each event.

1. Introduction. Our problem may be represented schematically as follows: Consider two pulse waves (Fig 1) of periods T_1, T_2, pulse widths t_1, t_2, and phases ϕ_1, ϕ_2. It is desired that these pulses overlap at least once within a given time interval; moreover, an overlap is not satisfactory unless its duration is at least as great as some assigned t_m. The starting phases ϕ_1 and ϕ_2 are unknown for *both* waves. Our problem, then, would appear to be to calculate as a function of time the probability of at least one overlap of duration at least t_m.

This probability will be calculated later, and, while mathematically adequate, is totally useless for practical application. This rather unusual occurrence in applied mathematics arises from sources generally kept in mind only by experimental physicists. Namely, the very nature of the science of measurement, involving as it always does at some stage, the use of the human senses, precludes

[1] This work was done in part under Contract No. OEMsr-411 between Harvard University and the Office of Scientific Research and Development, which assumes no responsibility for the accuracy of the statements contained herein.

the availability of mathematically exact values of the parameters of the problem. In other words, although experimental error can sometimes be made amazingly small, it can never be eliminated.

Now, as might be expected from the possibility that the waves may "lock in step", our probability is extremely erratic with respect to very minute changes in the periods T_1, T_2. For example, let $T_1 = T_2 = 100t_1 = 100t_2$ ($t_m = 0$); a simple direct calculation then shows that, for all times greater than $T_1 = T_2$, the desired probability is 0.03. Now if we let $T_1 = T_2 + \epsilon$, one wave will "creep up" on the other, and eventually (for times greater than T_1T_2/ϵ) the probability is unity! Thus it may very well happen in a practical application that the parameters are known to an accuracy essentially sufficient only to give the obvious result: $0 \le P \le 1$.

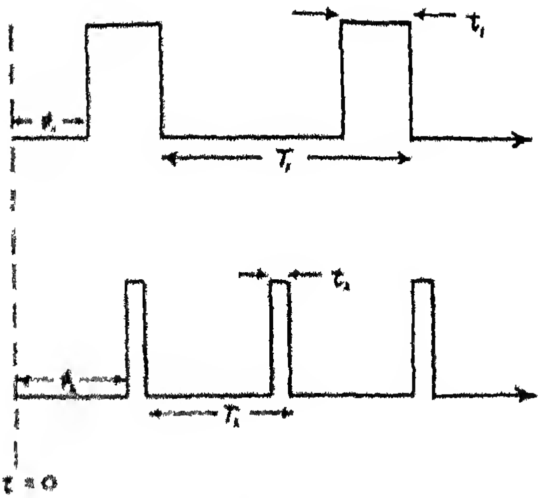


Fig. 1

In the practical problem originally considered, uncertainty in the data arose not only from experimental error but also from slight instability of equipment. Thus some means of averaging over variations in the periods had to be found if the analysis was to be of any practical value whatsoever.

For reasons which will appear in the later analysis, this smoothing entails difficulties which the author was unable to overcome with any great success; the nature of the results which have been obtained is discussed in the next section. These results involve several approximations which, generally speaking, are based on the assumption that the ratios t_i/T_i are both small.

It might be noted finally that the obviously favorable situations $t_1 > T_2$ or $t_2 > T_1$ often cannot be used because of numerous practical difficulties.

2. Results. In this section, we shall summarize the results of the later analysis for the benefit of those readers not interested in the latter. At the end of this section, there is an outline of the practical application of the formulas.

We shall continue to use the notation already introduced:

(1)
$$t_1, t_2 = \text{durations of the events};$$
$$T_1, T_2 = \text{periods of the events},$$
$$t_m = \text{minimum satisfactory duration of coincidence; and}$$
$$P = \text{probability of at least one satisfactory coincidence.}$$

We shall also use the (at present) rather arbitrary notation:

(2)
$$t = (\text{time} - t_m)$$
$$P_0 = (t_1 - t_m)(t_2 - t_m)/T_1T_2$$
$$w = (t_1 + t_2 - 2t_m)/T_1T_2.$$

The probability function for short time intervals is:

(3) $$P = P_0 + wt, \qquad \text{for } t \leq \text{Max}(T_1, T_2).$$

In any case:

(4) $$P \leq P_0 + wt.$$

As already explained, the functional dependence of P for large t is of no practical use due to its extremely erratic variation with small changes in the periods T_1, T_2.

For reasons which will later become apparent, the only type of averaging which has yet been carried to completion is the following. Consider that many trials of equal length are made and that in each individual trial, all the parameters are, by some mysterious device, held constant with absolute, mathematical exactitude. Assume for definiteness that $T_2 \leq T_1$. Between different trials, let t_2 and T_2 vary in such a way that T_1/T_2 takes all values within a range of $\frac{1}{3}$ with equal probability. (In the original problem, the ratios t_i/T_i necessarily remained constant.) The quantity f given below then represents that fraction of the trials in which the rigorous probability is *less* than an assigned value $= P_0 + Q$. Thus the *smaller* f is, the greater are the chances of success.

It must be admitted that this method assumes several things which are not true in practice. First, the parameters of the problem probably vary by at least a percent even within a single trial. More serious, the required variation in T_1/T_2 may, in the extreme case $T_1 = T_2$, demand as much as 33% variation in T_2. While considerable variation does occur, it is doubtful that it attains this magnitude. Finally, the method assumes that T_1 stays fixed as T_2 varies, whereas actually T_1 and T_2 vary simultaneously.

Despite these drawbacks, it was felt that the results were meaningful for the practical problem. In any case, they must serve until a more adequate analysis can be carried through.

The reader will notice that the final results have the form of a "probability of a

probability." It would thus seem that a simple integration would yield a true probability, but, unfortunately, the formulas for f are reasonably accurate only for $Q \leq \frac{1}{2}$. The final formula for f = fraction of trials in which $P < P_0 + Q$ is:

$$
(5)\quad f = \begin{cases} 1 & \text{for } tw < Q, \\ 1.216\,Q\left\{1 + \left(\dfrac{tw}{Q} - 1\right)\log\left(1 - \dfrac{Q}{tw}\right)\right\}, & \text{for } tw > Q, \quad Q \leq 1/2. \end{cases}
$$

This expression is subject to error from several sources. First it is an approximation to a number theoretic formula given in (31); this approximation is best for t and Q/w large compared to $\mathrm{Max}(T_1, T_2)$. A *completely general* comparison of (31), and (5) = (33) is given in Fig. 2, where the agreement will be seen to be quite adequate even for relatively small t and Q/w. (The dotted contours are straight lines passing through the origin.) When t and Q/w are small this first source of error can be eliminated by using the solid contours of Fig. 2 in place of (5).

Secondly, formula (31) itself is an approximation and involves the use of simplified probability formulas and an assumption that P_0 and w are constant as T_2 varies. The maximum possible magnitude of these errors in (31) is given by (parentheses indicate functional dependence):

$$
(6)\qquad f(t\bar{w}, Q - p_0 - q) \leq f(tw, Q) \leq f(t\underline{w}, Q + p_0 + q),
$$

where, as T_2 varies,

$$\underline{w}, \bar{w} = \text{minimum, maximum values of } w$$

$$p_0 = \text{change in } P_0$$

$$q = \text{maximum value of } w^2T_1T_2.$$

Generally speaking, these errors are small if t_i/T_i are small and if t is large compared to $\mathrm{Max}(T_1, T_2)$. Also, there is considerable possibility that certain errors will cancel in such a way as to make (6) correct with $q = 0$.

We shall now outline the practical use of these results. Given nominal values of the parameters defined in (1), choose a convenient value for $Q \leq \frac{1}{2}$ (usually $Q = \frac{1}{2}$), and substitute into (2) to find tw/Q. From (5), one may then determine f = fraction of trials in which $P < P_0 + Q$. (Low values of f are thus desirable.) For computational convenience, (5) has been plotted in Fig. 3, while, above the range of Fig. 3, the following lies within 1% of (5).

$$
(7)\qquad f = 0.608(Q^2/tw) \qquad \text{for} \qquad tw > 10Q.
$$

Note also that (4) may often be of considerable use in quickly eliminating cases of very poor probability, and recall also that (3) will give the true, directly meaningful probability whenever t is no greater than $\mathrm{Max}(T_1, T_2)$.

Evaluation of the maximum possible error in f as so obtained is more complicated. If t and Q/w are small, Fig. 2 may be used to eliminate inexactness

due to the approximation of (31) by (5) = (33). Otherwise, this error may safely be assumed to be negligible (less than 0.025; (31) may be employed directly, but this is laborious unless Q/w is small). The remaining errors, given by (6), may change depending on how T_2 is assumed to vary. To make these bounds as close as possible, it is best to choose $T_2 = \mathrm{Min}(T_1, T_2)$ and then let

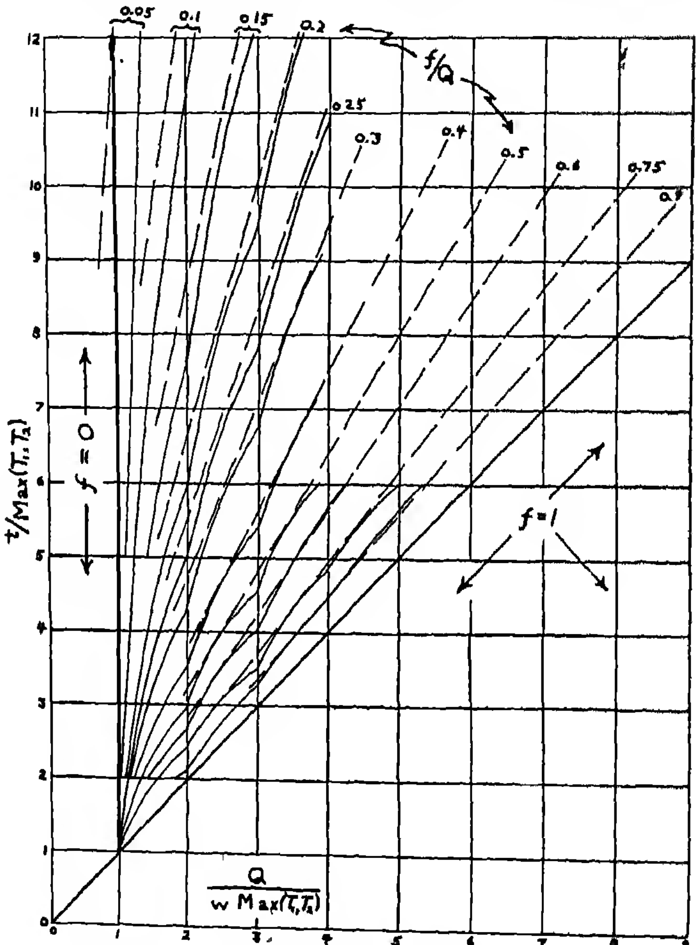


FIG. 2 Contours of f/Q —— (31); — — (33)

T_2 decrease from its nominal value by an amount sufficient to cause T_1/T_2 to increase by $\frac{1}{2}$

The reader may have noticed that f has a jump discontinuity as t passes through the value Q/w This is not the result of approximations; it occurs also in the number-theoretic formula (excepting only when $\mathrm{Max}(T_1, T_2) = \frac{1}{2}w$ and $Q = \frac{1}{2}$) and merely means that the "lock in" phenomena are suddenly able to have an effect when t becomes greater than Q/w.

3. The probability function. Our problem has already been represented by the pulse waves of Fig. 1. The starting phases ϕ_1, ϕ_2 of the waves are random, and we desire the probability P of at least one overlap of duration at least t_m within a given time interval. Manifestly $P = 0$ until time t_m; hence we shall give t the meaning already assigned in (2).

Consider any sub-interval of width t_m. The range of phases favorable to satisfactory coincidence on this interval is easily seen to be a rectangle with sides $(t_1 - t_m)$, $(t_2 - t_m)$ in the phase plane (ϕ_1, ϕ_2). By proper choice of the (arbitrary) zero-phase reference, the small rectangle favorable to coincidence on $(0, t_m)$ can be made to fall in the lower left corner of the phase plane (Fig. 4).

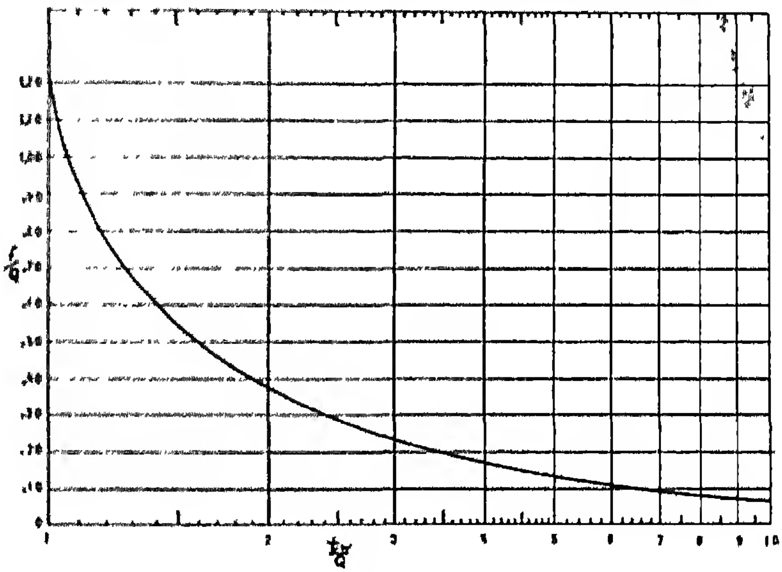


FIG. 3

As we allow the sub-interval (width t_m) to advance in time, this small rectangle will sweep out along a 45° line (Fig. 4); its horizontal displacement = vert. disp. is given by t as defined in (2). Since the phases must be measured modulo the periods, we must "switch back" the strip whenever it begins to leave the large rectangle: $0 \leq \phi_1 \leq T_1$, $0 \leq \phi_2 \leq T_2$; this is illustrated in Fig. 5.

The desired probability is then the area covered at least once by the strip divided by (T_1T_2), the total available area of the phase plane.

Using Fig. 4, one can easily show that, before the strip begins to overlap itself:

$$P = P_0 + wt, \tag{8}$$

where t, P_0, w are defined in (2).

A rectangle with opposite sides identified, as in Fig. 5, is topologically equivalent to a torus. This gives a good geometric picture of the overlap phenomena.

The strip winds diagonally about the torus until eventually (in general after several full circuits) it strikes sufficiently near its starting point to overlap itself on one edge It then begins to fill the chinks between the previous circuits, and this single overlap continues until the chinks are almost filled. The strip then approaches its starting point from the side opposite to that on which single overlap occurred. Thereafter, only the center section of the strip is effective in increasing the area covered. This double overlap continues until the entire torus has been covered. A degenerate case is possible in which the strip, upon its first overlap, begins to retrace exactly its former path and the torus is never fully covered This corresponds to interlocking of the original waves of Fig. 1.

A rigorous proof of the above statements may be constructed by using the fact that each change in behavior can occur only at the starting point. In this manner, it is easily shown that. (a) single and double overlap occur in that order,

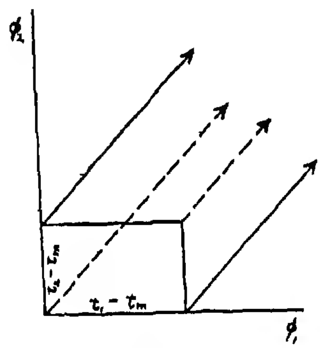


FIG 4

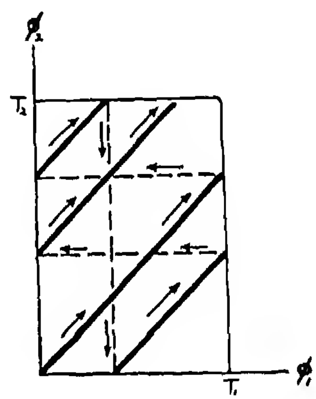


FIG 5

(b) the strip area effective in covering changes only upon a change in the type of overlap, and (c) the two types of overlap must occur on opposite sides of the starting point.

The facts (a, b, c) may then be used to derive the probability function. For the analytic analysis, it is best to return to the (ϕ_1, ϕ_2) plane. Overlap of any type will first occur when the "unswitched-back" strip approaches sufficiently near a point (n_1T_1, n_2T_2) where n_1 and n_2 are non-negative integers not both zero. The analysis is greatly shortened by noticing that the behavior is completely determined by the distance of the line $\phi_1 = \phi_2$ from such points (even though the strip is not centered on this line), while the width of the strip is (Fig. 4) $wT_1T_2/\sqrt{2}$.

A slight fine-structure may arise in the probability function where it changes slope, depending on whether or not the leading corner of the moving rectangle strikes one of the sides of the original small rectangle. These effects are small if t_i/T_i are small and will be neglected below by supposing the strip to be gen-

erated by a line segment oriented perpendicularly to its path. The error arising from this procedure consists essentially in a delay or advance in the time at which P changes slope. It may be seen that the maximum effect represents a delay of $\Delta t = wT_1T_2/2$. The error introduced is then less than $\Delta t\sqrt{2}$ multiplied by that portion of the total width of the strip which becomes ineffective due to the overlap considered. The sum of these effects must be less than that given by using the total width of the strip; this gives the maximum error $w^2T_1T_2/2$.

The results of the method outlined are then as follows. Single overlap occurs at $t = s$ where

$$s = \tfrac{1}{2}(m_1T_1 + m_2T_2), \tag{9}$$

and (m_1, m_2) is that pair of non-negative integers not both zero such that s is a minimum and

$$p_1 = \left|\frac{m_1}{T_2} - \frac{m_2}{T_1}\right| < w. \tag{10}$$

Double overlap occurs at $t = d$, where

$$d = \tfrac{1}{2}(n_1T_1 + n_2T_2), \tag{11}$$

and (n_1, n_2) is that pair of non-negative integers not both zero such that d is a minimum and the conditions

$$\begin{aligned} &\left|\frac{n_1}{T_2} - \frac{n_2}{T_1}\right| < w, \\ &\left(\frac{n_1}{T_2} - \frac{n_2}{T_1}\right)\left(\frac{m_1}{T_2} - \frac{m_2}{T_1}\right) < 0, \end{aligned} \tag{12}$$

are satisfied. If we set

$$p_2 = p_1 + \left|\frac{n_1}{T_2} - \frac{n_2}{T_1}\right| - w, \tag{13}$$

the probability function is then

$$P \begin{cases} = P_0 + wt & \text{for } t \leq s, \\ = P_0 + sw + (t - s)p_1 & \text{for } s \leq t \leq d, \\ = P_0 + sw + (d - s)p_1 + (t - d)p_2 & \text{for } d \leq t, \end{cases} \tag{14}$$

where it is understood that $P = 1$ if (14) gives $P > 1$.

The degenerate case where the waves interlock is given correctly by this formalism. Namely, if the strip starts to retrace its path exactly, then $p_1 = 0$ and the second part of (12) shows that d does not exist. Equation (14) then gives the correct result: P rises to the value $P_0 + sw$ and never increases further.

4. The method of smoothing. We have already discussed in section 1 the inadequacy of the formal mathematical solution (14) for purposes of practical

application Either mathematical analysis or intuitive consideration of interlock shows that the erratic behavior of P is due almost entirely to small changes in the *ratio* T_1/T_2. As this ratio passes through certain rational values, possibilities of interlock appear and disappear. Consequently, we next alter (14) to a form in which the dependence on this ratio is more evident.

We may, without loss of generality, assume:

$$T_1 = 1, \qquad T_2 < 1. \tag{15}$$

Also introduce the standard notation:

$$[x] = (\text{largest integer} \leq x). \tag{16}$$

It will then be seen that (10) and (12) may be thrown into the form:[2]

(17) k = smallest positive integer such that $p_1 = |\,ke - i\,| < w$ (i = integer);

(18) K = smallest positive integer such that $|\,Ke - I\,| < w$ and also

$$(ke - i)\,(Ke - I) < 0 \ (I = \text{integer}),$$

where either

$$e = \frac{1}{T_2} - \left[\frac{1}{T_2}\right], \qquad \text{or} \qquad e = 1 + \left[\frac{1}{T_2}\right] - \frac{1}{T_2}. \tag{19}$$

Now from (9) and (10), we note that s differs from m_1T_1 by at most $wT_1T_2/2$, while from (11) and (12), d differs from n_1T_1 by less than the same amount. Moreover, by the second half of (12), d is thereby made too small if s has been made too large and vice versa. Hence the use of these approximations in (14) will contribute an error certainly less than $w^2T_1T_2/2$. Adding the error discussed in section 3, the total introduced thus far cannot exceed $w^2T_1T_2$.

We thus use in the present notation $s = k$, $d = K$; (13) and (14) then become:

$$p_2 = p_1 + |\,Ke - I\,| - w \tag{20}$$

(21)
(a) $P = P_0 + wt,$ for $t \leq k$

(b) $P = P_0 + kw + (t - k)p_1,$ for $k \leq t \leq K$

(c) $P = P_0 + kw + (K - k)p_1 + (t - K)p_2,$ for $K \leq t$

where, as before, $P = 1$ if (21) gives a value greater than unity. Equations (17)–(21) are the formulation which will be used, with conditions (15), henceforth.

We wish now to smooth P with respect to variations in e. The number-theoretic requirement (17) is extremely difficult to work with. For reasons of simplicity, then, we shall assume that e is the only parameter which changes as

---

[2] Note that, even though the periods appear explicitly only in (19) hereafter, all the following equations are true only for $T_1 = 1$ (This is evident if we recall that w has the dimensions of inverse time) Thus we are definitely assuming that T_1 = constant.

T_2 is varied. The errors which may arise from this assumption are treated at the end of section 5.

From (19) or from the absolute value signs in (17), (18) it will be seen that all possible situations arise if e varies merely from zero to one-half. In order that this should entail as little variation in T_2 as possible, our conventions should be chosen as already stated in (15). Even under these circumstances, a maximum variation of 33% in T_2 may be required to cover the range $e = 0$ to $\frac{1}{2}$.

Equation (21) cannot be used directly without the interpretational convention there noted. This leads to difficulties of treatment which the author was unable to solve. The difficulties may be avoided by the following device, which admittedly has less direct significance than an averaged value for P.

We enquire after the fraction f of the range of e over which P has a value (at fixed t) less than some given value $Q + P_0$. We may then say that, if a large number of trials each of length t is made, then in f of them, the probability of coincidence will be less than $Q + P_0$.

5. Calculation of f. The exceptional behavior of P is that caused by interlock possibilities. This corresponds to $p_1 = 0$ in (17). Thus the exceptional values of P center about the points $e = i/k$, where i and k are relatively prime (otherwise, k would not be the smallest integer satisfying (17)). Moreover, by a standard theorem [1], $k \leq 1/w$. Thus the critical points form the Farey series of order $1/w$ in the range $(0, \frac{1}{2})$. About each Farey point, we may suspect that there will be an interval over which k is constant, and that the entire range may thereby be divided up into ranges of constant k.

In thinking about the use of (17) in a typical calculation, it is convenient to eliminate the integer i by representing multiples of e as a series of points progressing around and around a circle of unit circumference. When $e = i/k$, the kth multiple will (after i revolutions) coincide with the origin; this and the earlier points, it is easily shown, will be distributed uniformly about the circle with a separation $1/k$.

As e moves away from the Farey point, k will, by definition (17), remain constant until either (a) the point ke moves a distance greater than w from the origin or (b) an *earlier* point moves to a distance less than w from the origin (Fig. 6).

Let (me) be that *earlier* point nearest (initially $1/k$ from) the origin and moving toward it as e varies in a particular direction. Of course,

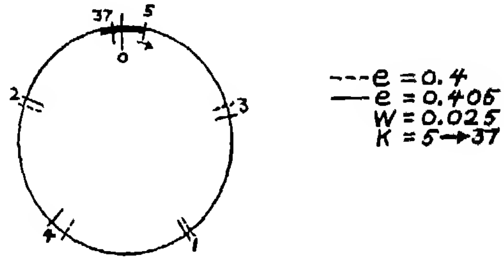
$$m < k. \tag{22}$$

For each Farey point, there will be two values of m; one for decreasing e and one for increasing e. If we introduce the new variable: h = the absolute value of the change in e from the Farey point i/k, then each point, ne, on the reference circle will move a distance nh, and (17) gives as the conditions for constant k (Fig. 7):

$$\begin{aligned} &\text{(a) } w > kh = p_1, \\ &\text{(b) } mh < (1/k) - w. \end{aligned} \tag{23}$$

Thus we have divided the range $(0, \frac{1}{2})$ into small ranges where k (and m) are fixed. The number of small ranges is roughly twice the number of Farey points in $(0, \frac{1}{2})$.

Within each small range p_1, K, p_2 still vary with e. The behavior of p_1 is



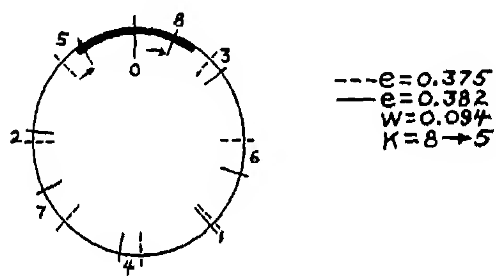


FIG. 6

already given in (23a); we shall find that we do not need p_2. Using (18) and Fig. 7, it may easily be shown that:

(24) $$K = m + jk + k,$$

where

(25) $$j + a = (1 - mkh - kw)/k^2h, \qquad j = [j], \qquad 0 \leq a < 1.$$

From (23a), (24), (25), we obtain:

(26) $$(K - k)p_1 = 1 - kw - ak^2h \qquad (0 \leq a < 1).$$

Having thus divided the range of e into small regions within each of which the number-theoretic requirements (17, 18) take a relatively simple form, we must now turn to the calculation of f = that fraction of the range $e = (0, \frac{1}{2})$ over which $P < P_0 + Q$ at fixed t. We shall specialize the further analysis to the case $Q \leq \frac{1}{4}$. This considerably shortens the discussion and yields essentially all the useful results of the more general inquiry.

We first note from (21) that, since $p_2 < p_1 < w$ (i.e. because of (4)), we have $P < P_0 + Q$ independently of e if $t < Q/w$

(27) $$f = 1, \qquad \text{for} \qquad t < Q/w.$$

Similar reasoning shows on the other hand that, when $t > Q/w$, those regions with $k \geq Q/w$ do not contribute to f. In the following, we shall therefore employ:

(28) $$k < Q/w < t, \qquad Q \leq \tfrac{1}{2}.$$

Equation (28) implies that we must use either (21b) or (21c); we shall next show that we do not need (21c). The value of P whenever (21c) is applicable is certainly greater than $(P_0 + kw + (K - k)p_1)$. From (26), this value is equal to $(P_0 + 1 - ak^2h)$. Now from (28), $w < 1/2k$, whence by (23a) $h < 1/2k^2 < 1/2ak^2$ (since $a < 1$). Thus $(P_0 + 1 - ak^2h) > P_0 + \frac{1}{2} \geq P_0 + Q$, and consequently (21c) never applies until $P \geq P_0 + Q$. (This means merely that the double overlap discussed in section 3 cannot occur until at least half the torus is covered.) Accordingly, we can confine our attention entirely to (21b) in any further discussion of f.

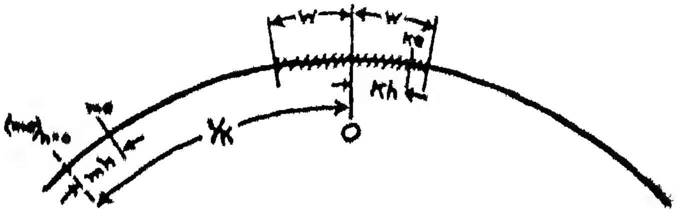


FIG. 7

Substituting for p_1 from (23) and recalling that $(t - k)$ is positive (by (28)), we find from (21b) that the condition $P < P_0 + Q$ becomes:

(29) $$h < \frac{Q - kw}{k(t - k)}.$$

However, h is subject also to the restrictions (23), which insure that we do not stray from the small region where k is constant. We assert that (29) implies (23) and may therefore be used as the final expression of the requirement $P < P_0 + Q$.

To prove this, note first that (29) and (28) immediately give $h < w/k$, which is (23a). Secondly, (28) implies $1/k > 2w$ so that, using (23a) and (22): $(1/k) - w > w > kh > mh$, which is (23b).

Thus we arrive at the result that f receives contributions only from those elementary regions where k satisfies (28) and that the contribution of each such region is governed by (29).

Since the variable h was defined as the absolute value of the change of e from the Farey point i/k, each Farey point (satisfying (28)) contributes an amount equal to twice[3] the right-hand side of (29). Since this amount is independent

[3] This is not true of the Farey points 0 and $\frac{1}{2}$, the ends of the range of e, but the terms $k = 1, 2$ in (31) correctly account for these contributions since $\phi(1) = \phi(2) = 1$.

of i, we may immediately sum over all Farey points i/k with fixed k. There are $\frac{1}{2}\phi(k)$ such points[3] in the range $(0, \frac{1}{2})$, where Euler's function ϕ is defined by:

(30) $\phi(k)$ = the number of integers $\leq k$ and relatively prime to k.

(Note that $\phi(k)$ is even for $k \geq 3$ since if k and i have no common divisor > 1, neither do k and $k - i$.)

Thus, summing over all these contributions and dividing by the length of the total range:

$$f = 2 \sum_{1 \leq k < Q/w} \phi(k) \frac{Q - kw}{k(t - k)}, \qquad \text{for } t > Q/w. \tag{31}$$

Regarding error in (31) due to the inaccuracy of (21), note that this can enter only when we set $P = P_0 + Q$ in deriving (29). Actually the difference between (21b) and the correct value of P will change as e is changed so that there is considerable possibility that these effects will cancel out in (31). (In fact, a detailed study shows that the error in (21b) assumes opposite signs as e varies in opposite directions from any given Farey point.) In any case, because (31) is monotone in Q, the error in (31) can be no greater than that found by substituting $Q \pm w^2T_1T_2$ for Q. Taking account also of the variation of P_0 with T_2, the same argument establishes the "Q-dependence" of (6) given in section 2.

Finally, we investigate the error due to change in w with T_2. If $\overline{w}$ is the maximum value of w, Farey points with $k < Q/\overline{w}$ are certain to contribute to f, and this contribution will be at least as great as $(Q - k\overline{w})/k(t - k)$ so that $f > f(\overline{w})$. On the other hand, if $\underline{w}$ is the minimum value of w, Farey points with $k \geq Q/\underline{w}$ cannot possibly contribute to f, and the remaining points can contribute no more than $(Q - k\underline{w})/k(t - k)$ so that $f < f(\underline{w})$. Hence we arrive at the final statement (6) in section 2

6. Approximations for f. Computational difficulties in the use of (31) suggested approximating it by a more readily computed expression. By a standard theorem [1, p. 266]:

$$\phi(k) \approx 6k/\pi^2 \tag{32}$$

We may then approximate (31) by:

$$\begin{aligned} f &= 1.216 \int_{\frac{1}{2}}^{(Q/w)+\frac{1}{2}} \frac{Q - kw}{t - k}\, dk \\ &= 1.216\, Q \left(1 + \frac{tw - Q}{Q} \log \frac{tw - Q - \frac{1}{2}w}{tw - \frac{1}{2}w}\right). \end{aligned}$$

If Q/w is large compared to $\frac{1}{2}$ (recall $t > Q/w$), this becomes very nearly:

$$f = 1.216\, Q \left(1 + \left(\frac{tw}{Q} - 1\right) \log \left(1 - \frac{Q}{tw}\right)\right) \qquad \text{for } t > Q/\underline{w}. \tag{33}$$

Despite the cavalier derivation of (33), its agreement with (31) is remarkably

close. Fig. 2 shows a perfectly general comparison of (31) and (33), where the agreement will be seen to be fairly good even for t and Q/w of the order of 4 or 5. Note also that (33) nearly always gives a value of f that is too large.

For completeness, we may repeat (27).

$$f = 1 \qquad \text{for} \qquad t < Q/w. \tag{34}$$

Note that only the dimensionless quantities tw, Q enter into (33, 34) which are therefore independent of the normalization (15).

REFERENCE

[1] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Clarendon Press, Oxford, 1938, p. 30.

NONPARAMETRIC ESTIMATION, III. STATISTICALLY EQUIVALENT BLOCKS AND MULTIVARIATE TOLERANCE REGIONS—THE DISCONTINUOUS CASE

By John W. Tukey

Princeton University

1. Summary. In Paper II of this series [2, 1947] it was shown that if n functions and a sample of n were used to divide the population space into $n + 1$ blocks in a particular way, and *if the joint cumulative of the functions were continuous*, then the $n + 1$ fractions of the population, corresponding to the $n + 1$ blocks, were distributed symmetrically and simply.

In Paper I of this series [1, 1945] it was shown that the one-dimensional theory of tolerance regions could be extended to the discontinuous case, if equalities were replaced by inequalities.

In this paper the results of Paper II will be extended to the discontinuous case with the same weakening of the conclusion. The devices involved are more complex, but the nature of the results is the same (See Section 5).

As a tool, it is shown that any n-variate distribution can be represented in terms of an n-variate distribution with a continuous joint cumulative (in fact, with uniform univariate marginals), where each variate of the given distribution is a different monotone function of the corresponding variate from the continuous distribution.

2. Introduction. The importance of extending the simple results of the continuous case to the more complex results of the discontinuous case may not be clear at first thought. Yet all the data with which the statistician actually works comes from *discontinuous distributions*. Often these distributions are very fine-grained—the distributions of the number of eggs laid by codfish and of the measured wavelengths of a spectral line (measured in 0.000001 Å) do not have *large* concentrated probabilities, but *all* their probability *is* concentrated at discrete points. Insofar as the considerations of the theoretical statistician apply to the data as received rather than to the "data" of a more or less imaginary model, these considerations apply to data with a discrete distribution. When his theories are erected on a basis of a probability density function, or even a continuous cumulative, there is a definite extrapolation from theory to practice. It is, ultimately, a responsibility of the mathematical statistician to study discrete models and find out the dangerous large effects and the pleasant small effects which go with such extrapolation. *We all deal with discrete data, and must sooner or later face this fact.*

In order to deal with the discontinuous case, we must face two problems: (we assume that the reader is familiar with Paper II [2])

(1) What to do about "ties"?

(2) Finite probabilities associated with cuts.

The first of these is peculiar to the multivariate situation and can be easily explained by an example. Consider the three points in the plane with coordinates (1, 9), (3, 9) and (2, 6). Let the first two functions be y and x, then the procedure of Section 4 of Paper II [2] is not unique—two possibilities arise:

Alternative A. (1, 9) is selected as having the largest y, and (3, 9) as having the largest x among the remaining (two) points, hence $S_1 = \{(x, y)|y > 9\}$, $S_2 = \{(x, y)|y < 9, x > 3\}$, $S_{3;4} = \{(x, y)|y < 9, x < 3\}$.

Alternative B. (3, 9) is selected as having the largest x, and (2, 6) as having the largest x among the remaining (two) points, hence $S_1 = \{(x, y)|y > 9\}$, $S_2' = \{(x, y)|y < 9, x > 2\}$, $S_{3;4}' = \{(x, y)|y < 9, x < 2\}$.

Notice that $S_2' \neq S_2$. The procedure is not unique. In the continuous case, ties happen with probability zero, hence their consequences could be neglected. This is now no longer the case.

This difficulty is solved by using more functions and the idea of lexicographical (like a dictionary!) ordering. In the simplest case, we add no new functions and proceed as follows: If there is a unique i for which $\varphi_1(w_i)$ is maximal, select it. Otherwise look among the w_i for which $\varphi_1(w_i)$ is maximal—look at the values of $\varphi_2(w_i)$. If there is a unique such i for which $\varphi_2(w_i)$ is maximal, select it. If not, go on to $\varphi_3(w_i) \cdots$. This procedure leads to a specific i unless $\varphi_h(w_j) = \varphi_h(w_k)$ for h and some $j \neq k$. But in this case it does not matter whether j or k is selected, the set of m-tuples $(\varphi_1(w_i), \varphi_2(w_i), \cdots, \varphi_m(w_i))$ remaining will be the same, although the indices i will not. But the indices play no role in the actual construction.

As an example, consider the following 20 four-letter words as a sample and let there be four functions—φ_i being the negative of the position in the alphabet of the i-th letter of the word. (Thus $a > b > c > \cdots > z$.)

Sample: meet, west, made, gone, come, back, said, that, maid, well, with, with, just, week, very, near, edge, this, last, have. (*The Law of the Three Just Men*, Edgar Wallace, pp. 159–160).

Selections: *back*, *made*, near, (gone, come, edge, have. The fourth selection to be made at random among these four.) The inferences which can be made about the four-letter words in Edgar Wallace's writing vocabulary are left to the reader.

We have just given one rule for breaking ties, one which chooses Alternative B in our example. But we might prefer a rule which chooses Alternative A. To get more generality, we have only to take M functions, $M \geq m$, and let $\varphi_{p(1)}$, $[\varphi_{p(2)}, \cdots, \varphi_{p(m)}$, (where we may suppose $p(1) = 1$ without loss of generality) play the role just taken by $\varphi_1, \varphi_2, \cdots, \varphi_m$. Thus if the maximum of $\varphi_1(w)$ is not unique proceed to $\varphi_2(w)$, thence to $\varphi_3(w), \cdots$, thence to $\varphi_m(w)$. For the second block, start with $\varphi_{p(2)}$, then $\varphi_{p(2)+1}, \varphi_{p(2)+2}, \cdots, \varphi_m$. And so on. The choice

$$\varphi_1(x, y) = y,$$

$$\varphi_2(x, y) = -x,$$

$$\varphi_3(x, y) = xe^y,$$

$$\varphi_4(x, y) = x,$$

$$\varphi_5(x, y) = y^2,$$

with $p(1) = 1$ and $p(2) = 4$, leads to Alternative A above. (Note that φ_3 is a dummy in the sense that it is never used.) The problem of ties, which was a problem in uniqueness of construction, is thus dealt with.

Next we must deal with the cuts. When we made S_1, S_2 and $S_{2|4}$ in Alternative A, we omitted some points, namely

$$T_1 = \{(x, y) \mid y = 9\}, \quad \text{and} \quad T_2 = \{(x, y) \mid y < 9, x = 2\}.$$

In the continuous case this did not matter, since these sets had probability zero and could be avoided. Here they cannot, and we shall have to consider a family of blocks (in the wide sense) as consisting of the blocks S *and the cuts* T. The solution of the univariate case in Paper I [1] shows us that what we must expect is that:

$$\Pr \{ \text{coverage } S_i + T_{i-1} + T_i > t\} \geq \Pr \{ \text{coverage of one}$$

$$\text{continuous-case block} > t\} \geq \Pr \{ \text{coverage } S_i > t\}.$$

That is, if we want a certain set of blocks to cover (together) *at least* a certain amount with a certain probability we must add the adjoining cuts; and if we want a certain set of blocks to cover *at most* a certain amount with a certain probability we may add only these cuts which do not adjoin blocks not in our set. By introducing the cuts explicitly, we solve the second problem.

In order to reduce the size of the cuts, our detailed definitions will differ in detail from those which we have used so far. In the example, where the functions leading to Alternative A are used; we place in S_1 not only the points with $y > 9$, but also those with $y = 9$ and $-x > -1$; we place in S_2 not only the points with $y < 9$ and $x > 3$ and the points with $y < 9$, $x = 3$, $y^2 > 49$, but also those with $y = 9$ and $-x < -1$. Proceeding in this way, we reduce T_1 to the point $x = 1$, $y = 9$ and T_2 to the point $x = 3$, $y = 9$. This reduction can only diminish the probability associated with the cuts, but we cannot be sure that it will reduce it to zero.

Only in the quasi-trivial case, where the probability that all functions shall tie together is zero, do we return to the simplicity of the continuous case. This case is quasi-trivial because it does not arise with discrete probabilities, and real observations always involve discrete probabilities.

Having discussed the results, we should now briefly touch on the methods. The proof of the main theorems depends on two facts:

(1) a representation theorem, (5.3), and

(2) a lemma, (6.1) which shows that m functions would be enough if (i) the distribution were fixed, and (ii) cases of probability zero were neglected. The

representation theorem has been outlined in the summary. It is analogous to, but a definite extension of the one used in Paper I [1]. It seems to be new in statement, though not in thought—it will surprise few probability theorists. The novel element is the monotonicity of the functions, which is utterly essential for our purposes.

The lemma allows us to reduce the general case to the case of no extra functions, where the reduction must be made differently for each underlying distribution. The reduced functions are then represented by the representation theorem and the results of Paper II [2] are taken over. The results are stated in a form independent of the underlying distribution and the particular representation, hence they apply in general

The last paragraph stresses the principle common to Paper I [1] and this paper. It is natural to call it the "iceberg principle," and to sketch it as follows: "We have some information about the visible one-ninth of the iceberg, and we want to conclude something about this visible part. If we can imagine another eight-ninths, consistent with the part we know, and if using that we can prove something expressed solely in terms of the visible part, then this is the required proof. (The only essential is to be able to match *every* visible part.)" Both the reduced functions (which depend on the underlying distribution) and the uniform variables used to represent them are part of the invisible eight-ninths which "could be there."

3. Terminology and Notation. In general we use the terminology and notation of Paper II [2], and we shall continue to assume that all functions concerned in the argument are measurable.

Given two finite sequences of the same length, we write $(a_1, a_2, \cdots, a_m) > (b_1, b_2, \cdots, b_m)$ if *any of the following hold:*

$$a_1 > b_1,$$

$$a_1 = b_1, \text{ and } a_2 > b_2,$$

$$a_1 = b_1,\ a_2 = b_2, \text{ and } a_3 > b_3,$$

$$\cdots$$

$$a_i = b_i \text{ for } i < m, \text{ and } a_m > b_m .$$

This is the lexicographical order referred to above. (We interpret $(a_1, a_2, \cdots, a_m) < (b_1, b_2, \cdots, b_m)$ to mean $(b_1, b_2; \cdots, b_m) > (a_1, a_2, \cdots, a_m)$ and $=$ to mean identity.)

3.1 DEFINITION: *Given a sequence of real-valued functions* $\varphi_1, \varphi_2, \cdots, \varphi_M$ *and a sequence of starting indices* $p(1), p(2), \cdots, p(m)$, (which we shall often refer to, briefly, as an m-system of functions, $\varphi_1, \varphi_2, \cdots, \varphi_M$, without explicitly mentioning the starting indices), *the functions* $\Phi_1, \Phi_2, \Phi_3, \cdots, \Phi_m$ *are defined as follows:*

$$\Phi_k(w) = \{\varphi_{p(k)}(w), \varphi_{p(k)+1}(w), \cdots, \varphi_M(w)\}, \tag{3.2}$$

the values of Φ_k being sequences of $M - p(k) + 1$ numbers. (In these terms, the rule for tie-breaking already explained becomes "select an i for which $\Phi_k(w_i)$ is maximal (in the sense of lexicographical ordering)".)

4. The blocks and cuts determined by n points. 4. DEFINITION: *Given an m-system of functions $\varphi_1, \varphi_2, \cdots, \varphi_M$ and n points $w_1, w_2, \cdots, w_n$, $(m \leq n)$ the corresponding blocks and cuts are given by the following procedure:* (the Φ's are defined in 3.1) *First $i(1)$ is selected to maximize $\Phi_1(w_i)$, when*

$$S_1 = \{w \mid \Phi_1(w) > \Phi_1(w_{i(1)})\},$$

$$T_1 = \{w \mid \Phi_1(w) = \Phi_1(w_{i(1)})\}.$$

Next, $i(2)$ is selected $\neq i(1)$ and to maximize $\Phi_2(w_i)$ among such i, when

$$S_2 = \{w \mid \Phi_1(w) < \Phi_1(w_{i(1)}), \Phi_2(w) > \Phi_2(w_{i(2)})\},$$

$$T_2 = \{w \mid \Phi_1(w) < \Phi_1(w_{i(1)}), \Phi_2(w) = \Phi_2(w_{i(2)})\}.$$

$\cdots\cdots\ \cdots$(the construction is perfectly analogous to II-4.1)

$$S_{m|n+1} = \{w \mid \Phi_k(w) < \Phi_k(w_{i(k)}), k = 1, 2, \cdots, m\}.$$

4.2 DEFINITION: *If $m = n$, then $S_{n|n+1}$ is also denoted by S_{n+1}.*

If $m > n$, then only $\Phi_1, \Phi_2, \cdots, \Phi_n$ are used and $S_{n|n+1}$ is also denoted by S_{n+1}.

We denote by λ a subset (possibly none, possibly all) of the indices $1, 2, \cdots, m$ and $m|n + 1$ or, in case $m \geq n$ of the indices $1, 2, \cdots, n + 1$.

4.3 DEFINITION. *The block-group B_λ consists of the union of all S_i with i in λ and all T_i with both i and $i + 1$ in λ($m + 1$ means $m \mid n + 1$).*

The closed block-group $\bar{B}_\lambda$ consists of the union of all S_i with i in λ and all T_i with either i or $i + 1$ in λ

Given any set we define its *coverage* as the proportion of the population falling into it (here the underlying probability distribution appears for the first time in this section), and we use

4.4 DEFINITION: *The coverage of B_λ is denoted by $C(\lambda)$ and that of $\bar{B}_\lambda$ by $\bar{C}(\lambda)$.*

Thus, given a family of functions φ and n points w, the space of the w is divided into blocks and cuts, these are joined together into block-groups, and these block-groups have coverages. Thus, if the family of functions is fixed, the n points determine these coverages, and, if the points are chance points, the coverages are chance numbers.

5. Statement of results. Having discussed the construction, we can now state the results.

(5.1) THEOREM $A^*_{m|n+1}$. *Let $\varphi_1, \varphi_2, \cdots, \varphi_M$ be any m-system of functions and let $W_1, W_2, \cdots, W_n$, where $m \leq n$, be a sample from any distribution, let the blocks, cuts, block-groups and coverages be formed, as described above, using the*

same (unknown) *distribution for forming the coverages. Then, if* $\alpha_1, \alpha_2, \cdots, \alpha_p$ *are any set of* λ*'s (each* λ *is a set of indices!),*

$$\Pr\{C(\alpha_1) < a_1, C(\alpha_2) < a_2, \cdots, \bar{C}(\alpha_k) > a_k, \cdots, \bar{C}(\alpha_p) > a_p\}$$
$$\geq \Pr\{t(\alpha_1) < a_1, t(\alpha_2) < a_2, \cdots, t(\alpha_k) > a_k, \cdots, t(\alpha_p) > a_p\},$$

where $t(\lambda) = \Sigma t_i$ *for* i *in* λ, $t_{m|n+1} = t_{m+1} + \cdots + t_{n+1}$, *and* $t_1, t_2, \cdots, t_{n+1}$ *have a uniform distribution on the barycentric simplex.* (Compare Theorem $A_{m|n+1}$ of Paper II [2].)

In particular,

$$\Pr\{C(i) < a\} \geq I_a(1, n) \geq \Pr\{\bar{C}(i) < a\}, \quad i = 1, 2, \cdots, m,$$

where $I_a(1, n)$ *is the incomplete Beta-function.*

(5.2) THEOREM B^*_{n+1}. *Let* $\varphi_1, \varphi_2, \cdots, \varphi_M$ *be any* n-*system of functions and let* $W_1, W_2, \cdots, W_n$ *be a sample from any distribution. Then*

$$\Pr\{C(\alpha_1) < a_1, C(\alpha_2) < a_2, \cdots, \bar{C}(\alpha_k) > a_k, \cdots, \bar{C}(a_p) > a_p\}$$
$$\geq \Pr\{t(\alpha_1) < a_1, t(\alpha_2) < a_2, \cdots, t(\alpha_k) > a_k, \cdots, t(\alpha_p) > a_p\},$$

where $t(\lambda) = \Sigma t_i$ *for* i *in* λ *and* $t_1, t_2, \cdots, t_{n+1}$ *have a uniform distribution on the barycentric simplex. In particular,*

$$\Pr\{C(i) < a\} \geq I_a(1, n) \geq \Pr\{\bar{C}(i) < a\}, \quad i = 1, 2, \cdots, n+1.$$

For convenience of reference, we also state the representation theorem as:

(5.3) THEOREM C. *Let* $X_1, X_2, \cdots, X_n$ *have any joint* n-*variate distribution. Then there exist (real) functions* $g_1, g_2, \cdots, g_n$ *and a joint distribution for* $U_1, U_2, \cdots, U_n$ *such that,*

(i) *the marginal distribution of each* U_i *is uniform on* [0, 1],

(ii) *each function* g *is non-decreasing,*

(iii) *the distribution of* $g_1(U_1), g_2(U_2), \cdots, g_n(U_n)$ *is identical with that of* $X_1, X_2, \cdots, X_n$.

6. The functions ψ. The aim of this section is to prove

(6.1) LEMMA. *Given any* m-*system of functions* $\varphi_1, \varphi_2, \cdots, \varphi_M$, *there exist real functions* $\psi_1, \psi_2, \cdots, \psi_M$ *such that,* if $W_1, W_2, \cdots, W_n$ *are a sample from the distribution concerned:*

(6.2) $\Pr\{\psi_i(W_j) = \psi_i(W_k)$, *but* $\psi_{i+h}(W_j) \neq \psi_{i+h}(W_k)$ *for some* $h > 0\} = 0$.

(6.3) $\Pr\{\Phi_i(W_j)$ *has a different relation to* $\Phi_i(W_k)$ *than that of* $\psi_i(W_j)$ *to* $\psi_i(W_k)\} = 0$, *where by relation is meant* $>$, $=$, *or* $<$.

The ψ_i will depend on the underlying probability distribution. Thus they are useful in the proof, but could not replace the Φ_i in the statement of the theorems.

(6.4) LEMMA. *Let* $\Phi(w)$ *have its values in a totally ordered set,* (i.e. *always either* $\Phi_1 < \Phi_2$, $\Phi_1 = \Phi_2$ *or* $\Phi_1 > \Phi_2$) *and let* W *have a distribution. Consider the function* ψ,

$$\psi(w) = \Pr\{\Phi(W) < \Phi(w)\}.$$

Let $W_1, W_2, \cdots, W_n$ be a sample from the same distribution, then, with probability one, the relation ($<$, $=$, or $>$) *between $\Phi(W_j)$ and $\Phi(W_k)$ is the same as that between $\psi(W_j)$ and $\psi(W_k)$.*

If $\Phi(w_j) < \Phi(w_k)$, then $\psi(w_j) \leq \psi(w_k)$, if $\psi(w_j) < \psi(w_k)$, then $\Phi(w_j) < \Phi(w_k)$. These follow directly from the definition. To prove the lemma, then, we must show that

(i) $\psi(w_j) = \psi(w_k)$ but $\Phi(w_j) < \Phi(w_k)$ occurs with probability zero.

We may clearly assume that the totally ordered set is complete, and that, in particular, it contains the symbols $-\infty$ and $+\infty$. Consider the real function of an abstract variable,

$$F(s) = \Pr\{\Phi(W) < s\}.$$

It is a monotone function, with $F(-\infty) = 0$ and $F(+\infty) = 1$. We can therefore, given $\epsilon > 0$, select elements $-\infty = s_0 < s_1 < s_2 < \cdots < s_k = +\infty$ such that

$$0 \leq F(s_{i+1}) - F(s_i + 0) < \epsilon.$$

If (*i*) occurs, then $\Phi(w_j)$ and $\Phi(w_k)$ belong either to the same open interval (s_i, s_{i+1}) or one belongs to an open interval and the other is its upper endpoint. The probability of either of these happening is at most

$$\frac{n(n-1)}{2}\{F(s_{i+1}) - F(s_i + 0)\}^2 + n\{F(s_{i+1}) - F(s_i + 0)\}\{F(s_{i+1} + 0) - F(s_{i+1})\}.$$

Summing this over all intervals yields an estimate of

$$\frac{n(n-1)}{2} \operatorname*{Max}_i \{F(s_{i+1}) - F(s_i + 0)\} = \frac{n(n-1)}{2}\epsilon.$$

Since this goes to zero, the lemma is established.

We turn now to the proof of (6.1). The system of functions $\varphi_1, \varphi_2, \cdots, \varphi_M$ define the $\Phi_1, \Phi_2, \cdots, \Phi_m$ according to Section 3. These define $\psi_1, \psi_2, \cdots, \psi_m$ according to lemma (6.4) just proved. Applying this m times proves (6.3). Recalling that $\Phi_i(w_j) = \Phi_i(w_k)$ implies $\Phi_{i+h}(w_j) = \Phi_{i+h}(w_k)$, we see that (6.3) implies (6.2).

7. The notation $F(x + \lambda\cdot 0)$. All practitioners of analysis are familiar with $F(x + 0)$ and $F(x - 0)$, defined by

$$F(x \pm 0) = \lim_{h \downarrow 0} F(x \pm h).$$

We now generalize this formal notation to

$$F(x + \lambda\cdot 0) = \frac{1+\lambda}{2}F(x + 0) + \frac{1-\lambda}{2}F(x - 0), \tag{7.1}$$

where we will, in our immediate applications, need only λ's between -1 and $+1$

(although the definition applies in general). Notice, for example, that

$$F(x - 0) \leq F(x + \lambda \cdot 0) \leq F(x + 0), \quad \text{for } -1 \leq \lambda \leq 1,$$

that if F is continuous at x,

$$F(x + \lambda \cdot 0) = F(x \pm 0) = F(x),$$

that the condition for F to be normalized is

$$F(x + 0 \cdot 0) = F(x).$$

A similar definition is made for functions of two variables, namely

$$F(x + \lambda \cdot 0, y + \mu \cdot 0) = \frac{1 + \mu}{2} F(x + \lambda \cdot 0, y + 0) + \frac{1 - \mu}{2} F(x + \lambda\ 0, y - 0)$$

$$= \frac{1 + \lambda}{2} F(x + 0, y + \mu \cdot 0) + \frac{1 - \lambda}{2} F(x - 0, y + \mu \cdot 0),$$

where the two right-hand sides are equal if, as is the case for cumulatives, all doubly one-sided limits exist.

If $F(x_1, x_2)$ is the joint cumulative of two variates, then, when all ordinates and abscissas involved are ordinates and abscissas of continuity,

$$Pr\{a \leq x \leq b, c \leq y < d\} = F(b, d) - F(b, c) - F(a, d) + F(a, c) \geq 0.$$

Passing to the limit in assorted ways, and taking linear combinations gives

$$\begin{aligned}(7.2) \qquad & F(b + \mu \cdot 0, d + \rho \cdot 0) - F(b + \mu \cdot 0, c + \nu \cdot 0) \\ & \qquad - F(a + \lambda \cdot 0, d + \rho \cdot 0) + F(a + \lambda \cdot 0, b + \nu \cdot 0) \geq 0,\end{aligned}$$

for $-\infty \leq a, b, c, d \leq +\infty$ and $-1 \leq \lambda, \mu, \nu, \rho \leq 1$. This will be of use shortly.

8. The representation theorem. It was shown in Paper I [1] of this series, that the uniform distribution on [0,1] could serve as the prototype of any variate —that is, that given a distribution, there is a monotone function g, so that $g(U)$ has the given distribution, where U has the uniform distribution on [0, 1]. (In Paper I, U was denoted by X^*).

In the notation of the last section, there is a function $\lambda(u)$, with $|\lambda(u)| \leq 1$, so that

$$(8.1) \qquad F(g(u) + \lambda(u) \cdot 0) = u,$$

for all u. (We may, and shall, require that $g(u) = -\infty$, for $u \leq 0$, and $g(u) = +\infty$ for $u \geq 1$). It is easy to see that $g(u)$ is unique except on a set of probability zero and that $\lambda(u)$ is unique (and in fact linear) on each open interval which contains no value of $F(x)$.

Each cumulative $F(x)$, then serves to define $g(u)$ and $\lambda(u)$ by the equation

(8.1). Two or more *independent* variates can be thrown back on a set of *independent* uniform variates by applying this process to their cumulatives separately.

Our present problem is to prove Theorem C (5.3), which applies to variates $X_1, X_2, \cdots, X_n$ which need not be independent. Let $F_i(x_i)$ be the (marginal cumulative of X_i, and use (8.1) to define $g_i(u_i)$ and $\lambda_i(u_i)$. Then define the joint distribution of $U_1, U_2, \cdots, U_n$ by

$$G(u_1, u_2, \cdots, u_n) = F(g_1(u_1) + \lambda_1(u_1)\cdot 0, \cdots, g_n(u_n) + \lambda_n(u_n)\cdot 0),$$

where $F(x_1, x_2, \cdots, x_n)$ is the joint cumulative of the $X_1, X_2, \cdots, X_n$.

We shall verify that this is the desired distribution in the case $n = 2$, leaving the general case to the reader. Consider $G(u_1, +\infty) = G(u_1, 1) = F(g_1(u_1) + \lambda_1(u_1)\cdot 0, +\infty)$. This is a cumulative, and so is $G(+\infty, u_2)$. In fact, using (8.1) they are each the uniform cumulative

$$G(u) = \begin{cases} 0, u \le 0, \\ u, 0 \le u \le 1, \\ 1, 1 \le u. \end{cases}$$

By (7.2) all second differences are positive, and hence $G(u_1, u_2)$ is a joint cumulative Since its marginals are uniform, it is continuous.

Finally,

$$Pr\{g_1(U_1) < s_1, g_2(U_2) < s_2\} = G(F(s_1 - 0, +\infty), F(+\infty, s_2 - 0))$$
$$= F(s_1 - 0, s_2 - 0),$$

since $g_1(u_1) < s_1$ is equivalent to $u_1 < F(s_1 - 0, +\infty)$ and $g_2(u_2) < s_2$ is equivalent to $u_2 < F(+\infty, s_2 - 0)$. Thus $g_1(U_1)$ and $g_2(U_2)$, have the given bivariate distribution.

9. Proof of main theorems. We come now to the proof of Theorems $A^*_{m|n+1}$ and B^*_{n+1}, and we begin with $A^*_{m|n+1}$. According to Lemma (6.1), the various indices, $i(1), i(2), \ldots, i(m)$ selected to determine the blocks will be the same, excluding cases of probability zero, whether the Φ_i or the ψ_i are used. Consider the first block, which takes the forms:

$$S_1' = \{W \mid \Phi_1(W) > \Phi_1(w_{i(1)})\}.$$
$$S_1'' = \{W \mid \psi_1(W) > \psi_1(w_{i(1)})\}.$$

Another application of Lemma (6.1) shows that these sets differ by a set of probability zero, and hence their coverages are identical. It will thus suffice to prove theorem $A^*_{m|n+1}$ for a fixed underlying distribution and the corresponding $\psi_1, \psi_2, \cdots, \psi_m$.

According to Theorem C (5.3), the m-variate distribution of the $\psi_i(W)$ can be represented in terms of uniformly distributed variates $U_1, \cdots, U_m$ and monotone functions $g_1(U_1), \cdots, g_m(U_m)$. Now $U_1, U_2, \cdots, U_m$ have a continuous joint

cumulative, so that theorem $A_{m|n+1}$ applies to a sample of n drawn from this m-variate population, with the coordinates themselves as the m functions. We shall denote the coordinates of the i-th element of this sample by $u_1(i), \cdots, u_m(i)$. Consider the first block,

$$S_1 = \{(U_1, \cdots, U_m) \mid U_1 > u_1(i(1))\}.$$

Its image, $g(S) = \{(g_1(U_1), \cdots, g_m(U_m)) \mid U_1 > u_1(i(1))\}$ contains

$$S_1^* = \{(g_1(U_1), \cdots, g_m(U_m)) \mid g(U_1) > g(u_1(i(1)))\},$$

and is contained in the union of S_1^* and T_1^*, where

$$T_1^* = \{(g_1(U_1), \cdots, g_m(U_m)) \mid g(U_1) = g(u_1(i(1)))\}.$$

Thus the conclusions of Theorem $A_{m|n+1}^*$ hold for $S_1^*, T_1^*, \cdots, S_m^*, T_m^*, S_{m|n+1}^*$.

Now while Theorem $A_{m|n+1}^*$ mentions the underlying W's implicitly, careful study shows that they are not really involved; only the joint distribution of the φ_i, which in our present case are the ψ_i, matters. Since this is the same for the $\psi_i(W)$ and the $g_i(U_i)$, Theorem $A_{m|n+1}^*$ must hold for the ψ_i and the theorem is proved.

Theorem B_{n+1}^* is again a special case of Theorem $A_{m|n+1}^*$.

REFERENCES

[1] H. Scheffé and J. W. Tukey, "Nonparametric Estimation I. Validation of order statistics," *Annals of Math Stat.*, Vol. 16 (1945), pp. 187–192 (Also cited as Paper I).

[2] J. W. Tukey, "Nonparametric Estimation II Statistically equivalent blocks and multivariate tolerance regions The continuous case," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 529–539 (Also cited as Paper II).

ASYMPTOTIC PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATE OF AN UNKNOWN PARAMETER OF A DISCRETE STOCHASTIC PROCESS

By Abraham Wald

Columbia University

Summary. Asymptotic properties of maximum likelihood estimates have been studied so far mainly in the case of independent observations. In this paper the case of stochastically dependent observations is considered. It is shown that under certain restrictions on the joint probability distribution of the observations the maximum likelihood equation has at least one root which is a consistent estimate of the parameter θ to be estimated. Furthermore, any root of the maximum likelihood equation which is a consistent estimate of θ is shown to be asymptotically efficient. Since the maximum likelihood estimate is always a root of the maximum likelihood equation, consistency of the maximum likelihood estimate implies its asymptotic efficiency.

1. Introduction. Let $\{X_i\}$, $(i = 1, 2, \cdots, \text{ad. inf.})$, be a sequence of chance variables. It is assumed that for any positive integral value n the first n chance variables $X_1, \cdots, X_n$ admit a joint probability density function $p_n(x_1, \cdots, x_n, \theta)$ involving an unknown parameter θ. The consistency relations

$$\int_{-\infty}^{+\infty} p_{n+1}(x_1, \cdots, x_{n+1}, \theta)\, dx_{n+1} = p_n(x_1, \cdots, x_n, \theta) \tag{1.1}$$

are assumed to hold

In what follows, for any chance variable u the symbol $E(u \mid \theta)$ will denote the expected value of u when θ is the true parameter value.

Let $t_n(x_1, \cdots, x_n)$ be an unbiassed estimate of θ. Cramér [1] and Rao [2] have shown that under some weak regularity conditions on the distribution function $p_n(x_1, \cdots, x_n, \theta)$, the variance of t_n cannot fall short of the value

$$\frac{1}{c_n(\theta)} = \frac{1}{E\left[\left(\frac{\partial \log p_n}{\partial \theta}\right)^2 \middle| \theta\right]}. \tag{1.2}$$

Thus, for any unbiassed estimate t_n the variate $\sqrt{c_n(\theta)}(t_n - \theta)$ has mean value zero and variance $\geqq 1$. An estimate t_n is called efficient if $\sqrt{c_n(\theta)}(t_n - \theta)$ has mean value zero and variance 1.

A sequence $\{t_n\}$, $(n = 1, 2, \cdots, \text{ad. inf.})$, of estimates is said to be asymptotically efficient if the mean of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ is zero and the variance of $\sqrt{c_n(\theta)}$ $(t_n - \theta)$ is 1 in the limit as $n \to \infty$. In the literature usually the additional requirement is made that the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ be normal.

To make a distinction between the two cases when the condition concerning the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ is fulfilled or not, we shall say that $\{t_n\}$ is asymptotically efficient in the wide sense if it satisfies the conditions concerning the mean and the variance of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ If, in addition, the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ is normal, we shall say that $\{t_n\}$ is asymptotically efficient in the strict sense. Clearly, if $\{t_n\}$ is asymptotically efficient in the strict sense, it is also asymptotically efficient in the wide sense.

A word of clarification is needed as to the meaning of the conditions concerning the mean and variance of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ One interpretation would be that the requirement is that

$$\lim_{n=\infty} E[\sqrt{c_n(\theta)}\ (t_n - \theta) \mid \theta] = 0 \tag{1.3}$$

and

$$\lim_{n=\infty} E[c_n(\theta)\ (t_n - \theta)^2 \mid \theta] = 1. \tag{1.4}$$

Another interpretation would be that the requirement is that the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$, provided that the limit distribution exists as $n \to \infty$, should have zero mean and unit variance. These two interpretations are certainly not equivalent. It seems to the author that the mean and variance of the limiting distribution is more relevant than the limits of the mean and the variance. We shall, therefore, adopt the following definition of asymptotic efficiency:

Definition: A sequence $\{t_n\}$ of estimates is said to be asymptotically efficient in the wide sense if a sequence $\{u_n\}$, $(n = 1, 2, \cdots$, ad. inf.), of chance variables exists such that

$$\lim_{n=\infty} E(u_n \mid \theta) = 0, \qquad \lim_{n=\infty} E(u_n^2 \mid \theta) = 1 \tag{1.5}$$

and

$$\sqrt{c_n(\theta)}(t_n - \theta) - u_n \tag{1.6}$$

converges stochastically to zero as $n \to \infty$. If, in addition, the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ exists and is normal, $\{t_n\}$ is said to be asymptotically efficient in the strict sense.

The reason that a sequence $\{u_n\}$ of chance variables is considered in the above definition, instead of the limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$, is that the existence of a limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ is not postulated. If a limiting distribution of $\sqrt{c_n(\theta)}\ (t_n - \theta)$ exists and if this limiting distribution has zero mean and unit variance, a sequence $\{u_n\}$ of chance variables satisfying the conditions (1.5) and (1.6) always exists. This can be seen as follows: Let T_n denote the chance variable $\sqrt{c_n(\theta)}\ (t_n - \theta)$ and let $F_n(t) = \text{prob.}\ \{T_n < t\}$. If a limit-

ing distribution of T_n exists and if this limiting distribution has zero mean and unit variance, then

$$(1.7)\qquad \lim_{a=\infty}\left[\lim_{n=\infty}\int_{-a}^{a} t\,dF_n(t)\right]=0 \qquad\text{and}\qquad \lim_{a=\infty}\left[\lim_{n=\infty}\int_{-a}^{a} t^2\,dF_n(t)\right]=1.$$

From (1.7) it follows that there exists a sequence $\{a_n\}$, $(n = 1, 2, \cdots, \text{ad. inf.})$, of positive values such that the following conditions are fulfilled:

$$(1.8)\qquad \lim_{n=\infty}\int_{-a_n}^{a_n} t\,dF_n(t)=0;\ \lim_{n=\infty}\int_{-a_n}^{a_n} t^2\,dF_n(t)=1;\ \lim_{n=\infty}\text{Prob}\,\{|T_n|>a_n\}=0.$$

Let u_n be a chance variable which is equal to T_n whenever $|T_n| \leqq a_n$, and equal to zero otherwise. Clearly, the sequence $\{u_n\}$ will satisfy conditions (1.5) and (1.6).

In the following section we shall formulate some assumptions concerning the probability density function $p_n(x_1, \cdots, x_n, \theta)$. It will then be shown in section 3 that there exists a root of the maximum likelihood equation

$$(1.9)\qquad \frac{\partial \log p_n}{\partial\theta}=0$$

which is asymptotically efficient at least in the wide sense.

2. Assumptions concerning the probability density $p_n(x_1, \cdots, x_n, \theta)$. We shall assume that there exists a finite non-degenerate interval A on the θ-axis such that the following conditions hold:

Condition 1. The derivatives $\frac{\partial^i p_n}{\partial\theta^i}$, $(i = 1, 2, 3)$, exist for all θ in A and for all samples $(x_1, \cdots, x_n)$ except perhaps for a set of measure zero. We have furthermore,

$$(2.1)\qquad \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} \underset{\theta\in A}{\text{l.u.b.}}\left|\frac{\partial^i p_n}{\partial\theta^i}\right| dx_1\cdots dx_n<\infty, \qquad (i=1,2).$$

Condition 2. For any θ in A we have $\lim_{n=\infty} c_n(\theta) = \infty$.

Condition 3. For any θ in A the standard deviation of $\frac{\partial^2 \log p_n}{\partial\theta^2}$ divided by the expected value of $\frac{\partial^2 \log p_n}{\partial\theta^2}$ (both computed under the assumption that θ is true) converges to zero as $n \to \infty$.

Condition 4. There exists a positive δ such that for any θ in A the expression

$$(2.2)\qquad \frac{1}{c_n(\theta)}\,E\left[\underset{\theta'}{\text{l.u.b.}}\left|\frac{\partial^3 \log p_n(x_1,\cdots,x_n,\theta')}{\partial\theta'^3}\right|\,\Big|\,\theta\right]$$

is a bounded function of n where θ' is restricted to the interval $|\theta' - \theta| \leqq \delta$.

In what follows in this section, as well as in section 3, the domain of θ will be

restricted to interior points of the interval A unless a statement to the contrary is explicitly made.

Clearly

$$(2.3)\qquad E\left(\frac{\partial \log p_n}{\partial \theta}\,\middle|\,\theta\right) = \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial p_n}{\partial\theta}\,dx_1\cdots dx_n\,.$$

It follows from Condition 1 that

$$(2.4)\qquad \int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial p_n}{\partial\theta}\,dx_1\cdots dx_n = \frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}p_n\,dx_1\cdots dx_n = 0.$$

Hence,

$$(2.5)\qquad E\left(\frac{\partial \log p_n}{\partial \theta}\,\middle|\,\theta\right) = 0.$$

We have

$$(2.6)\qquad \frac{\partial^2 \log p_n}{\partial\theta^2} = \frac{1}{p_n}\frac{\partial^2 p_n}{\partial\theta^2} - \left(\frac{\partial \log p_n}{\partial\theta}\right)^2.$$

Hence

$$(2.7)\qquad E\left(\frac{\partial^2 \log p_n}{\partial \theta^2}\,\middle|\,\theta\right) = E\left(\frac{1}{p_n}\frac{\partial^2 p_n}{\partial\theta^2}\,\middle|\,\theta\right) - c_n(\theta).$$

But

$$(2.8)\qquad E\left(\frac{1}{p_n}\frac{\partial^2 p_n}{\partial\theta^2}\,\middle|\,\theta\right) = 0,$$

because of Condition 1. From (2.7) and (2.8) we obtain

$$(2.9)\qquad E\left(\frac{\partial^2 \log p_n}{\partial \theta^2}\,\middle|\,\theta\right) = -c_n(\theta)$$

Conditions 3 and 4 will generally be fulfilled when the stochastic dependence of x_i on x_j decreases sufficiently fast with increasing value of $|i - j|$. For, in such cases, the following order relations will generally hold: The standard deviation of $\dfrac{\partial^2 \log p_n}{\partial\theta^2}$ will, in general, be of the order $\sqrt{n}$, the expected value of

$$\underset{|\theta'-\theta|\leq\delta}{\text{l.u.b.}}\left|\frac{\partial^3 \log p_n}{\partial\theta'^3}\right|$$

will usually be of the order n, and $\dfrac{c_n(\theta)}{n}$ will generally have a positive lower bound and a finite upper bound.

3. Proof that the maximum likelihood equation has a root which is an asymptotically efficient estimate of θ (at least in the wide sense). Let θ_0 denote the true parameter value and let θ be any other value. We put

$$\frac{\partial \log p_n}{\partial \theta} = \Phi_n, \qquad \frac{\partial^2 \log p_n}{\partial \theta^2} = \Phi_n' \qquad \text{and} \qquad \frac{\partial^3 \log p_n}{\partial \theta^3} = \Phi_n''. \tag{3.1}$$

Expanding $\Phi_n(x_1, \cdots, x_n, \theta)$ in a Taylor expansion around $\theta = \theta_0$ we obtain

$$\begin{aligned} \Phi_n(x_1, \cdots, x_n, \theta) &= \Phi_n(x_1, \cdots, x_n, \theta_0) + (\theta - \theta_0)\Phi_n'(x_1, \cdots, x_n, \theta_0) \\ &\quad + \tfrac{1}{2}(\theta - \theta_0)^2\Phi_n''(x_1, \cdots, x_n, \theta_n^*) \end{aligned} \tag{3.2}$$

where θ_n^* is some value between θ_0 and θ. Dividing both sides of (3.2) by $c_n(\theta_0)$ we obtain

$$\begin{aligned} \frac{\Phi_n(x_1, \cdots, x_n, \theta)}{c_n(\theta_0)} &= \frac{\Phi_n(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} \\ &\quad + (\theta - \theta_0)\frac{\Phi_n'(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} + \tfrac{1}{2}(\theta - \theta_0)^2\frac{\Phi''(x_1, \cdots, x_n, \theta_n^*)}{c_n(\theta_0)}. \end{aligned} \tag{3.3}$$

From Condition 3 and equation (2.9) it follows that

$$\operatorname*{plim}_{n=\infty} \frac{\Phi_n'(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} = -1 \tag{3.4}$$

where the operator plim stands for convergence in probability (stochastic convergence).

According to equation (2.5) the expected value of $\Phi_n(x_1, \cdots, x_n, \theta_0)$ is zero. Since the variance of $\Phi_n(x_1, \cdots, x_n, \theta_0)$ is equal to $c_n(\theta_0)$, and since $\lim_{n=\infty} c_n(\theta) = \infty$, we have

$$\operatorname*{plim}_{n=\infty} \frac{\Phi_n(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} = 0. \tag{3.5}$$

It follows from Condition 4 that for any θ with $|\theta - \theta_0| \leqq \delta$ we have

$$\frac{1}{c_n(\theta_0)} E(|\Phi_n''(x_1, \cdots, x_n, \theta_n^*|) = 0(1). \tag{3.6}$$

According to Markoff's inequality the probability that a positive random variable will exceed λ-times its expected value is not greater than $\frac{1}{\lambda}$. Hence, it follows from (3.6) that for any $\epsilon > 0$ we can find a positive value k_ϵ such that

$$\limsup_{n=\infty} \operatorname{Prob}\left\{\frac{1}{c_n(\theta_0)} |\Phi_n''(x_1, \cdots, x_n, \theta_n^*)| \geqq k_\epsilon\right\} \leqq \epsilon. \tag{3.7}$$

Let ρ be any given positive number. The probability that the maximum likelihood equation

$$\Phi_n(x_1, \cdots, x_n, \theta) = 0 \tag{3.8}$$

will have a root in the interval $(\theta_0 - \rho, \theta_0 + \rho)$ converges to one as $n \to \infty$. This follows easily from (3.3), (3.4), (3.5) and (3.7). Thus, we have shown that the maximum likelihood equation has a root $\bar{\theta}_n$ which is a consistent estimate, i e. it satisfies the relation

$$\text{plim}\,(\bar{\theta}_n - \theta_0) = 0. \tag{3.9}$$

We shall now show that if $\bar{\theta}_n$ is a root of the maximum likelihood equation (3.8) and if $\bar{\theta}_n$ is a consistent estimate, then $\bar{\theta}_n$ is also asymptotically efficient, at least in the wide sense. For this purpose we substitute $\bar{\theta}_n$ for θ in (3.3) and multiply both sides of the equation by $\sqrt{c_n(\theta_0)}$. We then obtain

$$0 = \frac{\Phi_n(x_1, \cdots, x_n, \theta_0)}{\sqrt{c_n(\theta_0)}} + \sqrt{c_n(\theta_0)}\,(\bar{\theta}_n - \theta_0)\,\frac{\Phi_n'(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} + \sqrt{c_n(\theta_0)}\,(\bar{\theta}_n - \theta_0)^2 v_n \tag{3.10}$$

where

$$v_n = \frac{1}{2}\,\frac{\Phi_n''(x_1, \cdots, x_n, \theta_n^*)}{c_n(\theta_0)}. \tag{3.11}$$

Let

$$y_n = \frac{\Phi_n(x_1, \cdots, x_n, \theta_0)}{\sqrt{c_n(\theta_0)}} \quad \text{and} \quad z_n = \sqrt{c_n(\theta_0)}\,(\bar{\theta}_n - \theta_0). \tag{3.12}$$

Then (3.10) given

$$-y_n = z_n\,\frac{\Phi_n'(x_1, \cdots, x_n, \theta_0)}{c_n(\theta_0)} + z_n(\bar{\theta}_n - \theta_0)\,v_n. \tag{3.13}$$

It follows from (3.7) and (3.9) that

$$\underset{n=\infty}{\text{plim}}\,(\bar{\theta}_n - \theta_0)\,v_n = 0. \tag{3.14}$$

From (3.4), (3.13) and (3.14) we obtain

$$-y_n = z_n(-1 + \xi_n) \tag{3.15}$$

where

$$\underset{n=\infty}{\text{plim}}\,\xi_n = 0. \tag{3.16}$$

Since $Ey_n = 0$ and $Ey_n^2 = 1$, it follows from (3.15) and (3.16) that

$$\underset{n=\infty}{\text{plim}}\,(z_n - y_n) = 0. \tag{3.17}$$

The asymptotic efficiency (in the wide sense) of $\bar{\theta}_n$ is an immediate consequence of (3.17). Our main result may be summarized in the following theorem:

THEOREM. *If the true value of the parameter θ is an interior point of an inter-*

val A satisfying the conditions 1 – 4, then the maximum likelihood equation (1.9) *has a root*[1] *which is a consistent estimate of θ. Furthermore, any root of* (1.9) *which is a consistent estimate of θ is also asymptotically efficient at least in the wide sense.*

Since the maximum likelihood estimate is a root of (1.9), it follows from the above theorem that whenever the maximum likelihood estimate is consistent, it is also asymptotically efficient at least in the wide sense.

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946.
[2] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters", *Bull. Calcutta Math. Soc.*, Vol. 37 (1945).

[1] The probability that (1.9) has at least one root converges to unity as $n \to \infty$.

DISTRIBUTION OF A ROOT OF A DETERMINANTAL EQUATION

By D. N. Nanda

Institute of Statistics, University of North Carolina

1. Summary. S. N. Roy [2] obtained in 1943 the distribution of the maximum, minimum and any intermediate one of the orots of certain determinantal equations based on covariance matrices of two samples on the null hypothesis of equal covariance matrices in the two populations. The present paper gives a different method of working out the distribution of any of these roots under the same hypothesis The distribution of the largest, smallest and any intermediate root when the roots are specified by their position in a monotonic arrangement has been derived for $p = 2, 3, 4$, and 5 by the new method. The method is applicable for obtaining the distribution of the roots of an equation of any order, when the distributions of the roots of lower order equations have been worked out

2. Introduction. If $x = \| x_{ij} \|$ and $x^* = \| x^*_{ij} \|$ are two p-variate sample matrices with n_1 and n_2 degrees of freedom respectively, and $S = xx'/n_1$ and $S^* = x^*x^{*\prime}/n_2$ are the covariance matrices which under the null hypothesis are independent estimates of the same population covariance matrix, then the joint distribution of the roots of the determinantal equation $|A - \theta(A + B)| = 0$ where $A = n_1S$ and $B = n_2S^*$ has been obtained by Hsu [1] in 1939. The distribution densty is

$$R(l, \mu, \nu) = \frac{\pi^{l/2} \cdot \prod_{i=1}^{l} \Gamma\left(\frac{l + \mu + \nu + i - 2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{\mu + i - 1}{2}\right) \Gamma\left(\frac{\nu + i - 1}{2}\right) \Gamma(i/2)} \tag{1}$$
$$\cdot \prod_{i=1}^{l} \theta_i^{(\mu/2)-1} \prod_{i=1}^{l} (1 - \theta_i)^{(\nu/2)-1} \prod_{i<j} (\theta_i - \theta_j),$$

$$(0 \leqq \theta_l \leqq \theta_{l-1} \leqq \cdots \theta_1 \leqq 1),$$

where $l = \min. (p, n_1)$, $\mu = |p - n_1| + 1$, and $\nu = n_2 - p + 1$.

This formula also gives the joint distribution of the squares of canonical correlations on the null hypothesis, that the two sets of variates are independent [1]. If

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1N} \\ x_{21} & \cdots & x_{2N} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ x_{p1} & \cdots & x_{pN} \end{bmatrix} \quad \text{and} \quad W = \begin{bmatrix} w_{11} & \cdots & w_{1N} \\ w_{21} & \cdots & w_{2N} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ w_{q1} & \cdots & w_{qN} \end{bmatrix}$$

are the observations on the two sets of canonical variates and the x's are normally distributed, independently of the w's, then the equation for the canonical roots is $| V_{xw}V_{ww}^{-1}V_{wx} - \theta V_{xx} | = 0$, where $\theta_i = r_i^2$ and $V_{xw} = XW'$ etc.. . . . It is observed that $V_{xw}V_{ww}^{-1}V_{wx}$ is like A with $n_1 = q$ and $V_{xx} - V_{xw}V_{ww}^{-1}V_{wx}$ is like B with $n_2 = N - q - 1$ and the above equation is reduced to the form $| A - \theta(A + B) | = 0$. It is under this condition that $R(l, \mu, \nu)$ gives the joint distribution density of $r_1^2, r_2^2, \cdots, r_l^2$, where $l = \min. (p, q)$, $\mu = | p - q | + 1$, and $\nu = N - p - q$.

3. Notation and preliminaries.

(a). Let

$$\prod_{i<j} (\theta_i - \theta_j) = \{1, 2, 3, \cdots, l\}.$$

It is known that the value of the Vandermonde determinant

$$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_l \\ \theta_1^2 & \theta_2^2 & \theta_3^2 & \cdots & \theta_l^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \theta_1^{l-1} & \theta_2^{l-1} & \theta_3^{l-1} & \cdots & \theta_l^{l-1} \end{vmatrix}$$

is equal to $\prod_{i>j} (\theta_i - \theta_j) = (-1)^l\{1, 2, 3, \cdots, l\}$.

Then

$$\begin{vmatrix} 1 & 1 & 1 \\ \theta_1 & \theta_2 & \theta_3 \\ \theta_1^2 & \theta_2^2 & \theta_3^2 \end{vmatrix} = (\theta_2 - \theta_1)(\theta_3 - \theta_2)(\theta_3 - \theta_1) = -\{1, 2, 3\},$$

but the determinant can also, by expansion in minors of the first row, be expressed as

$$-[\theta_1\theta_2\{1, 2\} + \theta_2\theta_3\{2, 3\} + \theta_3\theta_1\{3, 1\}]$$

where

$$\theta_1 - \theta_2 = \{1, 2\}.$$

Hence

(2) $$\{1, 2, 3\} = \theta_1\theta_2\{1, 2\} + \theta_3\theta_1\{3, 1\} + \theta_2\theta_3\{2, 3\}.$$

Similarly

(3) $$\{1, 2, 3, 4\} = \theta_1\theta_2\theta_3\{1, 2, 3\} - \theta_4\theta_1\theta_2\{4, 1, 2\} + \theta_3\theta_4\theta_1\{3, 4, 1\} - \theta_2\theta_3\theta_4\{2, 3, 4\},$$

and

$$\{1, 2, 3, 4, 5\} = \theta_1\theta_2\theta_3\theta_4\{1, 2, 3, 4\} + \theta_5\theta_1\theta_2\theta_3\{5, 1, 2, 3\} + \theta_4\theta_5\theta_1\theta_2\{4, 5, 1, 2\} + \theta_3\theta_4\theta_5\theta_1\{3, 4, 5, 1\} + \theta_2\theta_3\theta_4\theta_5\{2, 3, 4, 5\}. \tag{4}$$

It is seen that in the successive terms the θ's are present in a decreasing order.

(b). Let

$$(a, b; m, n) = y^m(1-y)^n \Big|_a^b = b^m(1-b)^n - a^m(1-a)^n,$$

and

$$(a, 1, b; m, n) = \int_a^b y^m(1-y)^n\, dy;$$

then

$$(a, 1, b; m+1, n) = -\frac{(a, b; m+1, n+1)}{m+n+2} + \frac{m+1}{m+n+2}(a, 1, b; m, n), \tag{4, a}$$

by a combination of the transformations obtained by partial integration and by breaking up $(1-y)^{n+1}$ into $(1-y)^n - y(1-y)^n$.

(c) Let

$$(a, 2, 1, b; m, n) = \int_{a<\theta_2<\theta_1<b} (\theta_1\theta_2)^m(1-\theta_1)^n(1-\theta_2)^n\{1, 2\}\, d\theta_1\, d\theta_2$$

$$(a, 2, b, 1, c; m, n) = \int_{a<\theta_2<b<\theta_1<c} (\theta_1\theta_2)^m(1-\theta_1)^n(1-\theta_2)^n\{1, 2\}\, d\theta_1\, d\theta_2,$$

and

$$(a, 3, b, 2, c, 1, d; m+1, n)$$
$$= \int_{a<\theta_3<b<\theta_2<c<\theta_1<d} (\theta_1\theta_2\theta_3)^{m+1}(1-\theta_1)^n(1-\theta_2)^n(1-\theta_3)^n\{1, 2, 3\}\, d\theta_1\, d\theta_2\, d\theta_3.$$

(d) Let

$$T_a^{b;m,n}\, g(y) = \int_a^b y^m(1-y)^n\, g(y)\, dy,$$

then

$$T_a^{b;m,n}(0, y; k, l) = (a, 1, b; m+k; n+l), \qquad (k > 0)$$

and

$$T_a^{b;m,n}(b, 1, c; k, l) = (a, 1, b; m, n)(b, 1, c; k, l).$$

With these preliminaries we proceed to derive the distribution of the roots.

4. Distribution of the largest root. Let us suppose that the roots are arranged in decreasing order such that for l roots we have

$$0 < \theta_l < \theta_{l-1} < \theta_{l-2}, \cdots, < \theta_2 < \theta_1 < 1.$$

If the distribution density $R(l, \mu, \nu)$ given by (1) be expressed as

$$R(l, m, n) = C(l, m, n) \prod_{i=1}^{l} \theta_i^m \prod_{i=1}^{l} (1 - \theta_i)^n \prod_{i<j} (\theta_i - \theta_j),$$

then the distribution of the largest root in the general case would be given by

$$Pr(\theta_1 \leqq x) = C(l, m, n)(0, l, l-1, \cdots, 2, 1, x; m, n).$$

Now we shall derive the distribution of the largest root for $l = 2, 3, 4,$ and 5.

(a) $l = 2$.

$$Pr(\theta_1 \leqq x) = C(2; m, n)(0, 2, 1, x; m, n).$$

$$\begin{aligned}(0, 2, 1, x; m, n) &= \int_{0<\theta_2<\theta_1<x} (\theta_1\theta_2)^m (1 - \theta_1)^n (1 - \theta_2)^n \{1, 2\}\, d\theta_1\, d\theta_2 \\ &= \int_{0<\theta_2<\theta_1<x} \theta_2^m (1 - \theta_2)^n \theta_1^m (1 - \theta_1)^n \{1, 2\}\, d\theta_1\, d\theta_2 \\ &= \int_{0<\theta_2<\theta_1<x} \theta_2^m (1 - \theta_2)^n \theta_1^{m+1} (1 - \theta_1)^n\, d\theta_1\, d\theta_2 \\ &\quad - \int_{0<\theta_1<\theta_2<x} \theta_2^m (1 - \theta_2)^n \theta_1^{m+1} (1 - \theta_1)^n\, d\theta_1\, d\theta_2 .\end{aligned}$$

The limits in the successive integrals are to be so adjusted as to keep the integrand same. Then using the notation given in section 3(*d*) and equation (4, *a*).

(5) $$(0, 2, 1, x; m, n) = T_0^{x;m,n}(y, 1, x; m + 1, n) - T_0^{x;m,n}(0, 1, y; m + 1, n)$$

or

$$\begin{aligned}(0, 2, 1, x; m, n) = T_0^{x;m,n}\Big[&-\frac{(y, x; m + 1, n + 1)}{m + n + 2} + \frac{m + 1}{m + n + 2}(y, 1, x; m, n) \\ &+ \frac{(0, y; m + 1, n + 1)}{m + n + 2} - \frac{(m + 1)}{m + n + 2}(0, 1, y; m, n)\Big].\end{aligned}$$

Now by a change in the order of integration,

$$T_0^{x,m,n}[(0, 1, y; m, n) - (y, 1, x; m, n)] = 0.$$

Therefore

$$\begin{aligned}(m + n + 2)(0, 2, 1, x; m, n) &= T_0^{x;m,n}[2(0, y; m + 1, n + 1) \\ &\quad - (0, y; m + 1, n + 1)] \\ &= 2(0, 1, x; 2m + 1, 2n + 1) \\ &\quad - (0, x; m + 1, n + 1)(0, 1, x; m, n).\end{aligned}$$

Hence

$Pr(\theta_1 \leqq x)$

$$= \frac{C(2, m, n)}{m + n + 2} [2(0, 1, x; 2m + 1, 2n + 1) - (0, x; m + 1, n + 1)(0, 1, x; m, n)]$$

$$= C(2, m, n) \left\{ \frac{2}{m + n + 2} \int_0^x y^{2m+1} (1 - y)^{2n+1} \, dy - \frac{x^{m+1}(1 - x)^{n+1}}{m + n + 2} \int_0^x y^m(1 - y)^n \, dy \right\}.$$

(b) $l = 3$. For this case we need certain results for $l = 2$ which can be easily obtained and are given below:

$$(6) \quad (a, 2, 1, b; m, n) = \frac{2}{m + n + 2} (a, 1, b; 2m + 1, 2n + 1) - \frac{1}{m + n + 2} [(0, a; m + 1, n + 1) + (0, b; m + 1, n + 1)] \times (a, 1, b; m, n)$$

and

$$(a, 2, b, 1, c; m, n) = \frac{1}{m + n + 2} [-(0, a; m + 1, n + 1)(b, 1, c; m, n) + (0, b; m + 1, n + 1)(a, 1, c; m, n) - (0, c; m + 1, n + 1)(a, 1, b; m, n)].$$

Now

$(0, 3, 2, 1, x; m, n)$

$$= \int_{0<\theta_3<\theta_2<\theta_1<x} (\theta_1\theta_2\theta_3)^m(1 - \theta_1)^n(1 - \theta_2)^n(1 - \theta_3)^n \{1, 2, 3\} \, d\theta_1 \, d\theta_2 \, d\theta_3$$

$$= \int_{0<\theta_3<\theta_2<\theta_1<x} (\theta_1 \theta_2 \theta_3)^m(1 - \theta_1)^n(1 - \theta_2)^n(1 - \theta_3)^n[\theta_1\theta_2\{1, 2\} + \theta_3\theta_1\{3, 1\} + \theta_2\theta_3\{2, 3\}] \, d\theta_1 \, d\theta_2 \, d\theta_3$$

(using equation (2))

$$= \int_{0<\theta_3<\theta_2<\theta_1<x} \theta_3^m(1 - \theta_3)^n(\theta_1\theta_2)^m(1 - \theta_1)^n(1 - \theta_2)^n \{1, 2\} \, d\theta_1 \, d\theta_2 + \int_{0<\theta_1<\theta_3<\theta_2<x} + \int_{0<\theta_2<\theta_1<\theta_3<x},$$

or

$$(0, 3, 2, 1, x; m, n) = T_0^{x;m,n}(y, 2, 1, x; m + 1, n) + T_0^{x;m,n}(0, 1, y, 2, x; m + 1, n) + T_0^{x;m,n}(0, 2, 1, y; m + 1, n),$$

but the θ's are to be always arranged in the same order, hence

$$\begin{aligned}(0, 3, 2, 1, x; m, n) = {} & T_0^{x;m,n}(y, 2, 1, x; m+1, n) \\ & - T_0^{x;m,n}(0, 2, y, 1, x; m+1, n) \\ & + T_0^{x;m,n}(0, 2, 1, y; m+1, n).\end{aligned}$$

Using equations (6) and (7), we have

$$\begin{aligned}&(0, 3, 2, 1, x; m, n) \\ &= \frac{T_0^{x,m,n}}{m+n+3}\{2(y, 1, x; 2m+3, 2n+1) - (y, 1, x; m+1, n) \\ &\quad \times [(0, y; m+2, n+1) + (0, x, m+2, n+1)] \\ &\quad - (0, 1, x; m+1, n)(0, y; m+2, n+1) + (0, 1, y; m+1, n)(0, x; m+2, n+1) \\ &\quad + 2(0, 1, y; 2m+3, 2n+1) - (0, 1, y; m+1, n)(0, y; m+2, n+1)\} \\ &= \frac{T_0^{x,m,n}}{m+n+3}\{2(y, 1, x; 2m+3, 2n+1) + (0, 1, y; 2m+3, 2n+1) \\ &\quad - (0, y, m+2, n+1)[(0, 1, x; m+1, n) + (0, 1, y; m+1, n) \\ &\quad + (y, 1, x; m+1, n)] - (0, x; m+2, n+1) \\ &\qquad [(y, 1, x; m+1, n) - (0, 1, y; m+1, n)]\} \\ &= \frac{T_0^{x,m,n}}{m+n+3}\{2(0, 1, x; 2m+3, 2n+1) \\ &\quad - 2(0, y; m+2, n+1)(0, 1, x; m+1, n) \\ &\quad - (0, x; m+2, n+1)[(y, 1, x; m+1, n) - (0, 1, y; m+1, n)]\}.\end{aligned}$$

Using equation (5), we have

$$\begin{aligned}(0, 3, 2, 1, x; m, n) = \frac{1}{(m+n+3)} & \{2(0, 1, x; 2m+3, 2n+1)(0, 1, x; m, n) \\ & - 2(0, 1, x; 2m+2, 2n+1)(0, 1, x; m+1, n) \\ & - (0, x; m+2, n+1)(0, 2, 1, x; m, n)\}.\end{aligned}$$

Hence

$$\begin{aligned}Pr(\theta_1 \leqq x) = \frac{C(3, m, n)}{(m+n+3)} & \{2(0, 1, x; 2m+3, 2n+1)(0, 1, x; m, n) \\ & - 2(0, 1, x; 2m+2, 2n+1)(0, 1, x; m+1, n) \\ & - (0, x; m+2, n+1)(0, 2, 1, x; m, n)\}.\end{aligned} \tag{8}$$

(c) $l = 4$. In order to determine $(0, 4, 3, 2, 1, x; m, n)$ we need the values of

$(a, 3, 2, 1, b; m, n)$, $(a, 3, b, 2, 1, c; m, n)$ and $(a, 3, 2, b, 1, c; m, n)$, which are obtained according to the procedure given above.

Now

$$\begin{aligned}
(0, 4, 3, 2, 1, x; m, n) &= \int_{0<\theta_4<\theta_3<\theta_2<\theta_1<x} \theta_4^m(1-\theta_4)^n(\theta_1\theta_2\theta_3)^m \\
&\qquad \cdot (1-\theta_1)^n(1-\theta_2)^n(1-\theta_3)^n\{1, 2, 3, 4\}\, d\theta_1\, d\theta_2\, d\theta_3\, d\theta_4 \\
&= \int_{0<\theta_4<\theta_3<\theta_2<\theta_1<x} \theta_4^m(1-\theta_4)^n(\theta_1\theta_2\theta_3)^m \\
&\qquad \cdot (1-\theta_1)^n(1-\theta_2)^n(1-\theta_3)^n[\theta_1\theta_2\theta_3\{1, 2, 3\} \\
&\qquad - \theta_4\theta_1\theta_2\{4, 1, 2\} + \theta_3\theta_4\theta_1\{3, 4, 1\} - \theta_2\theta_3\theta_4\{2, 3, 4\}]\, d\theta_1\, d\theta_2\, d\theta_3\, d\theta_4 \\
&= \int_{0<\theta_4<\theta_3<\theta_2<\theta_1<x} \theta_4^m(1-\theta_4)^n[(\theta_1\theta_2\theta_3)^{m+1} \\
&\qquad \cdot (1-\theta_1)^n(1-\theta_2)^n(1-\theta_3)^n]\{1, 2, 3\} \\
&\qquad - \int_{0<\theta_1<\theta_4<\theta_3<\theta_2<x} + \int_{0<\theta_2<\theta_1<\theta_4<\theta_3<x} - \int_{0<\theta_3<\theta_2<\theta_1<\theta_4<x} \\
&= T_0^{x,m,n}(y, 3, 2, 1, x; m+1, n) - T_0^{x,m,n}(0, 1, y, 3, 2, b; m+1, n) \\
&\quad + T_0^{x;m,n}(0, 2, 1, y, 3, x; m+1, n) - T_0^{x;m,n}(0, 3, 2, 1, y; m+1, n) \\
&= T_0^{x;m,n}(y, 3, 2, 1, x; m+1, n) - T_0^{x,m,n}(0, 3, y, 2, 1, b; m+1, n) \\
&\quad + T_0^{x;m,n}(0, 3, 2, y, 1, x; m+1, n) - T_0^{x;m,n}(0, 3, 2, 1, y; m+1, n).
\end{aligned}$$

Using the results of $(a, 3, 2, 1, b; m, n)$, $(a, 3, b, 2, 1, c; m, n)$ and $(a, 3, 2, b, 1, c; m, n)$, we have $Pr(\theta_1 \leqq x)$ equal to

$$\begin{aligned}
C(4, m, n)(0, 4, 3, 2, 1, x; m, n) &= \frac{C(4, m, n)}{m+n+4} \\
&\cdot \Big\{2(0, 1, x; 2m+5, 2n+1)(0, 2, 1, x; m, n) \\
&- \frac{2(0, 1, x; 2m+4, 2n+1)}{(m+n+3)}[2(0, 1, x; 2m+2, 2n+1) \\
&- (0, x; m+2, n+1)(0, 1, x; m, n) + (m+2)(0, 2, 1, x; m, n)] \\
&+ 2(0, 1, x; 2m+3, 2n+1)(0, 2, 1, x; m+1, n) \\
&- (0, x; m+3, n+1)(0, 3, 2, 1, x; m, n)\Big\}.
\end{aligned} \tag{9}$$

(d) $l = 5$. In the evaluation of the distribution of the largest root for $l = 5$; the following parts need to be calculated:

$$(a, 4, 3, 2, 1, b; m, n),\ (a, 4, b, 3, 2, 1, c; m, n),\ (a, 4, 3, b, 2, 1, c; m, n),$$
$$(a, 4, 3, 2, b, 1, c; m, n).$$

Proceeding along the lines indicated in the previous sections we get

$$Pr(\theta_1 \leq x) = \frac{C(5, m, n)}{(m+n+5)}\Big[2(0, 1, x; 2m+7, 2n+1)(0, 3, 2, 1, x; m, n)$$

$$- \frac{2(0, 1, x; 2m+6, 2n+1)}{(m+n+4)}\{2(0, 1, x; 2m+4, 2n+1)(0, 1, x; m, n)$$

$$- 2(0, 1, x; 2m+3, 2n+1)(0, 1, x; m+1, n)$$

$$- (0, x; m+3, n+1)(0, 2, 1, x; m, n)$$

$$+ (m+3)(0, 3, 2, 1, x; m, n)\} + \frac{2(0, 1, x; 2m+5, 2n+1)}{(m+n+4)}$$

$$(10) \qquad \cdot\Big\{2(0, 1, x; 2m+5, 2n+1)(0, 1, x; m, n)$$

$$- 2(0, 1, x; 2m+3, 2n+1)(0, 1, x; m+2, n)$$

$$- \frac{(0, x; m+3, n+1)}{(m+n+3)}[2(0, 1, x; 2m+2, 2n+1)$$

$$- (0, x; m+2, n+1)(0, 1, x; m, n)$$

$$+ (m+2)(0, 2, 1, x; m, n)]\Big\}$$

$$- 2(0, 3, 2, 1, x; m+1, n)(0, 1, x; 2m+4, 2n+1)$$

$$- (0, x; m+4, n+1)(0, 4, 3, 2, 1, x; m, n)\Big].$$

It is evident now that the above method can be used to derive the distribution for any value of l.

5. Distribution of the smallest root. Let $Pr[\theta_1 \leqq x/\mu, \nu] = P(x/\mu, \nu)$ where θ_1 is the largest root. Let us make the following transformations in the $R(l, \mu, \nu)$ distribution:

$$r_1 = 1 - \theta_l$$

$$r_2 = 1 - \theta_{l-1}$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$r_l = 1 - \theta_1;$$

then since $0 < \theta_l < \theta_{l-1} < \cdots < \theta_1 < 1$, we have $0 < r_l < r_{l-1} < r_{l-2} \cdots <$

$r_1 < 1$, and thus the domain of integration does not change. Hence the joint distribution of the r's can be expressed as

$$C(l, \nu, \mu) \prod_{i=1}^{l} (r_i)^{(\nu/2)-1} \prod_{i=1}^{l} (1 - r_i)^{(\mu/2)-1} \prod_{i<j} (r_i - r_j), \qquad 0 < r_l < \cdots < r_1 < 1.$$

Thus the r's have the same distribution as the θ's, but μ and ν are interchanged. Therefore

$$\begin{aligned} Pr(\theta_l \leq x) = Pr(1 - r_1 \leqq x) &= 1 - Pr(r_1 \leqq 1 - x) \\ &= 1 - P(1 - x/\nu, \mu). \end{aligned}$$

Hence, for getting the distribution of the smallest root, we have to change x into $1 - x$ and interchange m, n in the distributions of the largest roots and subtract the resultant probability from 1. The distributions for the smallest root are given below for $l = 2, 3, 4$ and 5.

(i) $l = 2$.

$$\begin{aligned} Pr(\theta_2 \leq x) &= 1 - Pr(\theta_1 \leqq 1 - x/n, m) \\ &= 1 - \frac{C(2, n, m)}{m + n + 2} \{2(0, 1, \overline{1 - x}, 2n + 1, 2m + 1) \\ &\qquad - (0, \overline{1 - x}, n + 1, m + 1)(0, 1, \overline{1 - x}, n, m)\}. \end{aligned} \tag{11}$$

(ii) $l = 3$.

$$\begin{aligned} P_r(\theta_3 \leqq x) = 1 - \frac{C(3, n, m)}{m + n + 3} \{&2(0, 1, \overline{1 - x}; 2n + 3, 2m + 1) \\ &\cdot(0, 1, \overline{1 - x}; n, m) \\ &- 2(0, 1, \overline{1 - x}; n + 1, m)(0, 1, \overline{1 - x}, 2n + 2, 2m + 1) \\ &- (0, \overline{1 - x}; n + 2, m + 1)(0, 2, 1, \overline{1 - x}; n, m)\}. \end{aligned} \tag{12}$$

(iii) $l = 4$.

$$\begin{aligned} Pr(\theta_4 \leqq x) = 1 - \frac{C(4, n, m)}{m + n + 4} \Big\{&2(0, 1, \overline{1 - x}, 2n + 5, 2m + 1) \\ &(0, 2, 1, \overline{1 - x}; n, m) \\ &- \frac{2(0, 1, \overline{1 - x}, 2n + 4, 2m + 1)}{(m + n + 3)} [2(0, 1, \overline{1 - x}; 2n + 2, 2m + 1) \\ &- (0, \overline{1 - x}; n + 2, m + 1)(0, 1, \overline{1 - x}; n, m) \\ &+ (n + 2)(0, 2, 1, \overline{1 - x}; n, m)] \\ &+ 2(0, 1, \overline{1 - x}; 2n + 3, 2m + 1)(0, 2, 1, \overline{1 - x}; n + 1, m) \\ &- (0, \overline{1 - x}, n + 3, m + 1)(0, 3, 2, 1, \overline{1 - x}; n, m)\Big\}. \end{aligned} \tag{13}$$

(iv) $l = 5$.

$$Pr(\theta_5 \leqq x) = 1 - \frac{C(5, n, m)}{(m+n+5)}\Big[2(0, 1, \overline{1-x}; 2n+7, 2m+1) \cdot (0, 3, 2, 1, \overline{1-x}; n, m)$$

$$- \frac{2(0, 1, \overline{1-x}; 2n+6, 2m+1)}{(m+n+4)}\{2(0, 1, \overline{1-x}; 2n+4, 2m+1) \cdot (0, 1, \overline{1-x}; n, m)$$

$$- 2(0, 1, \overline{1-x}; 2n+3, 2m+1)(0, 1, \overline{1-x}; n+1, m) - (0, \overline{1-x}; n+3, m+1)(0, 2, 1, \overline{1-x}; n, m)$$

$$+ (n+3)(0, 3, 2, 1, \overline{1-x}; n, m)\}$$

(14)
$$+ \frac{2(0, 1, \overline{1-x}; 2n+5, 2m+1)}{(m+n+4)}\Big\{2(0, 1, \overline{1-x}; 2n+5, 2m+1) (0, 1, \overline{1-x}; n, m)$$

$$-[2(0, 1, \overline{1-x}; 2n+3, 2m+1)(0, 1, \overline{1-x}; n+2, m)$$

$$- \frac{(0, \overline{1-x}; n+3, m+1)}{(m+n+3)}[2(0, 1, \overline{1-x}; 2n+2, 2m+1) - (0, \overline{1-x}; n+2, m+1)(0, 1, \overline{1-x}; n, m)$$

$$+ (n+2)(0, 2, 1, \overline{1-x}; n, m)]\Big\} - 2(0, 3, 2, 1, \overline{1-x}; n+1, m) \cdot (0, 1, \overline{1-x}; 2n+4, 2m+1)$$

$$- (0, \overline{1-x}; n+4, m+1)(0, 4, 3, 2, 1, \overline{1-x}; n, m)\Big].$$

6. Distribution of any intermediate root.

(i) $l = 3$.

$$Pr(\theta_2 \leqq x) = Pr(0 < \theta_3 < \theta_2 < \theta_1 < x) + Pr(0 < \theta_3 < \theta_2 < x < \theta_1)$$
$$= C(3, m, n)[(0, 3, 2, 1, x; m, n) + (0, 3, 2, x, 1; m, n)]$$

as the two probabilities are independent, or

$$Pr(\theta_2 \leqq x) = C(3, m, n)[(0, 3, 2, 1, x; m, n) + (0, 3, 2, x, 1, z; m, n)], \text{ where } z = 1$$

$$= \frac{C(3, m, n)}{m+n+3}\Big\{2(0, 1, x; 2m+3, 2n+1)(0, 1, x; m, n)$$

$$- 2(0, 1, x; m+1, n)(0, 1, x; 2m+2, 2n+1)$$

$$- (0, x; m+2, n+1)(0, 2, 1, x; m, n)$$

(15)
$$+ (x, 1, z, m, n)[2(0, 1, x; 2m+3, 2n+1)$$

$$- (0, x; m+2, n+1)(0, 1, x; m+1, n)]$$

$$- \frac{(x, z; m+2, n+1)}{m+n+2}[2(0, 1, x; 2m+1, 2n+1) - (0, x; m+1, n+1)(0, 1, x; m, n)]$$

$$+ (x, 1, z; m+1, n)(0, 1, x; m, n)(0, x; m+2, n+1)$$

$$- 2(x, 1, z; m+1, n)(0, 1, x; 2m+2, 2n+1)\Big\}.$$

(ii) $l = 4$.

$$
\begin{aligned}
Pr(\theta_2 \leqq x) &= Pr(0 < \theta_4 < \theta_3 < \theta_2 < \theta_1 < x; m, n) \\
&\quad + Pr\,(0 < \theta_4 < \theta_3 < \theta_2 < x < \theta_1\,; m, n) \\
&= C(4, m, n)[(0, 4, 3, 2, 1, x; m, n) + (0, 4, 3, 2, x, 1; m, n)]
\end{aligned}
$$

and

$$
\begin{aligned}
Pr(\theta_3 \leqq x) &= Pr(0 < \theta_4 < \theta_3 < \theta_2 < \theta_1 < x; m, n) \\
&\quad + Pr(0 < \theta_4 < \theta_3 < \theta_2 < x < \theta_1\,; m, n) \\
&\quad + Pr(0 < \theta_4 < \theta_3 < x < \theta_2 < \theta_1\,; m, n) \\
&= C(4, m, n)[(0, 4, 3, 2, 1, x; m, n) + (0, 4, 3, 2, x, 1; m, n) \\
&\quad + (0, 4, 3, x, 2, 1; m, n)].
\end{aligned}
$$

The different parts of these probabilities can be evaluated as indicated in section 4(d). Thus the method already indicated to obtain the distribution of the largest root also gives the distribution of any one of the roots.

7. Further problems. It is intended to prepare the probability distribution tables for small values of l. The results obtained in this paper are found to be useful in finding the distribution of the sum of the roots when the numbers of canonical variates in two sets differ by one. This problem is, however, being investigated further.

Acknowledgements. The author is highly indebted to Dr. P. L. Hsu for suggesting the problem and for guiding in this research, and is also thankful to Dr. Harold Hotelling for his suggestions and help in this work.

REFERENCES

[1] P. L. Hsu, "On the distribution of roots of certain determinantal equations", *Annals of Eugenics*, Vol. 3 (1939), pp. 250–258.

[2] S. N. Roy, "The individual sampling distribution of the maximum, the minimum and any intermediate of the 'p'-statistics on the null hypothesis", *Sankhyā*, December, 1943.

A k-SAMPLE SLIPPAGE TEST FOR AN EXTREME POPULATION

By Frederick Mosteller

Harvard University

1. Summary. A test is proposed for deciding whether one of k populations has slipped to the right of the rest, under the null hypothesis that all populations are continuous and identical. The procedure is to pick the sample with the largest observation, and to count the number of observations r in it which exceed all observations of all other samples. If all samples are of the same size n, n large, the probability of getting r or more such observations, when the null hypothesis is true, is about k^{1-r}

Some remarks are made about kinds of errors in testing hypotheses.

2. Introduction. The purpose of this paper is to describe a significance test connected with a statistical question called by the present author "the problem of the greatest one." Suppose there are several continuous populations $f(x - a_1)$, $f(x - a_2)$, $\cdots$, $f(x - a_k)$, which are identical except for rigid translations or slippages. Suppose further that the form of the populations and the values of the a_i are unknown. Then on the basis of samples from the k populations we may wish to test the hypothesis that some population has slipped further to the right, say, than any other. In other words, we may ask whether there exists an $a_i > \max(a_1, a_2, \cdots, a_{i-1}, a_{i+1}, \cdots, a_k)$. From the point of view of testing hypotheses, the existence of such an a_i is taken to be the alternative hypothesis. A significance test will depend also on the null hypothesis. We shall take as the null hypothesis the assumption that all the a's are equal: $a_1 = a_2 = \cdots = a_k$.

Using these assumptions it is possible to obtain parameter-free significance tests that some population has a larger location parameter (mean, median, quantile, say) than any of the other populations.

The problem of the greatest one is of considerable practical importance. Among several processes, techniques, or therapies of approximately equal cost, we often wish to pick out the best one as measured by some characteristic. Furthermore, we often wish to make a test of the significance of one of the methods against the others after noticing that on the basis of the sample values, a particular method seems to be best. The test provided in this paper allows an opportunity for inspection of the data before applying the test of significance.

The proposed test has the advantage of being rapid and easy to apply. However, the test is probably not very powerful, and in the form presented here, the test depends on having samples of the same size from each of the several populations. The equal-sample restriction is not essential to the technique, but since no very useful way of computing the significance levels for the unequal-sample case is known to the author, it does not seem worthwhile to give the formulas. They are easy to write down.

3. The test. Suppose we have k samples of size n each. It is desired to test the alternative hypothesis that one of the populations, from which the samples were drawn, has been rigidly translated to the right relative to the remaining populations. The null hypothesis is that all the populations have the same location parameter.

The test consists in arranging the observations in all the samples from greatest to least, and observing for the sample with the largest observation, the number of observations r which exceed all the observations in the $k - 1$ other samples. If $r \geq r_0$ we accept the hypothesis that the population whose sample contains the largest observation has slipped to the right of the rest and reject the null hypothesis that all the populations are identical; instead we accept the hypothesis that the sample with the largest observation came from the population with the rightmost location parameter. If $r < r_0$, we accept the null hypothesis.

The statements just made are not quite usual for accepting and rejecting hypotheses. Classically one would merely accept or reject the hypothesis that the a_i are all equal. The statements just made seem preferable for the present purpose.

Example. The following data arranged from least to greatest indicate the difference in log reaction times of an individual and a control group to three types of words on a word-association test. The differences in log reaction times have been multiplied by 100 for convenience. Longer reaction times for the individual are positive, shorter ones are negative. Does one type of word require a shorter reaction time for the individual relative to the control group than any other?

Concrete	Abstract	Emotional
−6	−16	−6
−6	−11	−5
−5	−3	−3
−5	−2	−2
−4	−2	−1
−3	−1	0
−1	−1	1
0	1	3
0	1	5
3	1	12
9	8	13
11	10	13
12	16	15
29	20	28

Here we have $k = 3$ samples of size $n = 14$ each! We note that the Abstract column has the most negative deviation, −16, and that there are two observations in that column which are less than all the observations in the other columns. Consequently $r = 2$. Under the null hypothesis the probability of ob-

taining 2 or more observations in one column less than all the observations in the others is about .33, so the null hypothesis is not rejected.

4. Derivation of test. Suppose we have k samples of size n, all drawn from the same continuous distribution function $f(x)$. Arranging observations within samples in order of magnitude the samples O_i are: $O_1 : x_{11}, x_{12}, \cdots, x_{1n}$; $O_2 : x_{21}, x_{22}, \cdots, x_{2n}$; $\cdots$; $O_k : x_{k1}, x_{k2}, \cdots, x_{kn}$.

If we consider some one sample O_i, separately, we can inquire about the probability that exactly r of its observations are greater than the greatest observation in the other $k - 1$ samples.

The total number of arrangements of the kn observations is

$$T = \frac{(kn)!}{(n!)^k}. \tag{1}$$

The number of ways of getting all n observations of O_i to be greater than all observations in the remaining samples is

$$N(n) = \frac{[(k-1)n]!}{(n!)^{k-1}0!}. \tag{2}$$

The number of ways of getting exactly $n - 1$ observations of O_i greater than all observations in the remaining samples is

$$N(n-1) = \frac{[(k-1)n+1]!}{(n!)^{k-1}1!} - \frac{[(k-1)n]!}{(n!)^{k-1}0!}. \tag{3}$$

More generally, the number of ways of getting exactly $r = n - u$ of O_i to be greater than all other observations in the remaining samples is

$$N(n-u) = \frac{[(k-1)n+u]!}{(n!)^{k-1}u!} - \frac{[(k-1)n+u-1]!}{(n!)^{k-1}(u-1)!}. \tag{4}$$

Therefore the number of ways of getting a run of $r = n - u$ *or more* observations in O_i greater than the rest is just

$$S(n-u) = \sum_{t=n-u}^{n} N(t) = \frac{[(k-1)n+u]!}{(n!)^{k-1}u!}. \tag{5}$$

However we do not choose our sample O_i at random or preassign it, as the demonstration has thus far supposed. Instead we choose that O_i which has the greatest observation in all the samples. This condition requires us to multiply $S(n-u)$ by the factor k. Consequently the probability that the sample with the largest observation has $r = n - u$ or more observations which exceed all observations in the other $k - 1$ samples is given by

$$P(r) = \frac{kS(r)}{T} - \frac{k(n!)}{(kn)!}\frac{(kn-r)!}{(n-r)!}. \tag{6}$$

As an incidental check we note in passing that

$$P(1) = \frac{k(n!)}{(kn)!}\frac{(kn-1)!}{(n-1)!} = \frac{kn}{kn} = 1.$$

We note that equation (6) may be rewritten as

$$P(r) = kC_{n-r}^{kn-r}/C_n^{kn}, \tag{7}$$

which is a useful form for some computations.

Table I gives the probability of observing r or more observations in the sample with the largest observation, among k samples of size n, which are more extreme in a preassigned direction than any of the observations in the remaining $k - 1$ samples.

5. Approximations. If we use Stirling's formula and approximations for $(1 + \alpha)^r$, for small values of α and r, we can write an approximation for equation (6) for large values of n with r and k fixed as follows

$$P(r) \sim \frac{1}{k^{r-1}}\left(1 - \frac{r(2r-1)(k-1)}{2kn}\right). \tag{8}$$

For very large n equation (8) yields

$$P(r) \sim \frac{1}{k^{r-1}}, \tag{9}$$

which is the value given in Table I for $n = \infty$ For many purposes the result given by equation (9) is quite adequate, as a glance at Table I will indicate.

6. Kinds of errors. In tests such as the one being considered here the classical two kinds of errors are not quite adequate to describe the situation.

As usual we may make the errors of

I) rejecting the null hypothesis when it is true,

II) accepting the null hypothesis when it is false.

But there is a third kind of error which is of interest because the present test of significance is tied up closely with the idea of making a correct decision about which distribution function has slipped furthest to the right. We may make the error of

III) correctly rejecting the null hypothesis for the wrong reason.

In other words it is possible for the null hypothesis to be false. It is also possible to reject the null hypothesis because some sample O_i has too many observations which are greater than all observations in the other samples. But the population from which some other sample say O_j is drawn is in fact the rightmost population In this case we have committed an error of the third kind.

When we come to the power of the test under consideration we shall compute the probability that we reject the null hypothesis because the rightmost population yields a sample with too many large observations. Thus by the power of

TABLE I

Probability of one of k samples of size n each having r or more observations larger than those of the other k − 1 samples

$k = 2$

n \ r	2	3	4	5	6
3	.400	.100			
5	.444	.167	.048	.008	
7	.462	.192	.070	.021	.005
10	.474	.211	.087	.033	.011
15	.483	.224	.100	.042	.017
20	.487	.231	.106	.047	.020
25	.490	.235	.110	.050	.022
∞	.500	.250	.125	.062	.031

$k = 3$

n \ r	2	3	4	5	6
3	.250	.036			
5	.286	.066	.011	.001	
7	.300	.079	.018	.003	.0004
10	.310	.089	.023	.005	.0011
15	.318	.096	.027	.007	.0018
20	.322	.100	.030	.009	.0023
25	.324	.102	.031	.009	.0026
∞	.333	.111	.037	.012	.0041

$k = 4$

n \ r	2	3	4	5	6
3	.182	.018			
5	.211	.035	.004	.0003	
7	.222	.043	.007	.0009	.0001
10	.231	.049	.009	.0015	.0002
15	.237	.053	.011	.0022	.0004
20	.241	.056	.012	.0026	.0005
25	.242	.057	.013	.0028	.0006
∞	.250	.062	.016	.0039	.0010

$k = 5$

n \ r	2	3	4	5
3	.143	.011		
5	.167	.022	.0020	.0001
7	.177	.027	.0033	.0003
10	.184	.031	.0046	.0006
15	.189	.034	.0056	.0008
20	.192	.035	.0062	.0010
25	.194	.036	.0065	.0011
∞	.200	.040	.0080	.0016

$k = 6$

n \ r	2	3	4	5
3	.118	.007		
5	.138	.015	.0011	.0000
7	.146	.018	.0019	.0001
10	.152	.021	.0026	.0003
15	.157	.023	.0032	.0004
20	.160	.024	.0035	.0005
25	.161	.025	.0037	.0005
∞	.167	.028	.0046	.0008

this test we shall mean the probability of both correct rejection and correct choice of rightmost population, when it exists.

Errors of the third kind happen in conventional tests of differences of means, but they are usually not considered although their existence is probably recognized. It seems to the author that there may be several reasons for this among which are 1) a preoccupation on the part of mathematical statisticians with the formal questions of acceptance and rejection of null hypotheses without adequate consideration of the implications of the error of the third kind for the practical experimenter, 2) the rarity with which an error of the third kind arises in the usual tests of significance.

In passing we note further that it is possible in the present problem for both the null hypothesis and the alternative hypothesis to be false when $k > 2$. This may happen when there are, say, two identical rightmost populations, and the remaining populations are shifted to the left. An examination of Table I will give us an idea of what will happen in such a case. If $k = 4$, we use $r = 3$ as about the .05 level. If two of the populations are slipped very far to the left, while the rightmost two populations are identical, in effect $k = 2$. In this case the probability of rejecting the null hypothesis is around .2. Consequently we accept the null hypothesis about 80 per cent of the time, and reject it 20 per cent of the time under these conditions. But neither hypothesis was true.

If we carry the discussion to its ultimate conclusion we would need a fourth kind of error for these troublesome situations. There are still other kinds of errors which will not be considered here.

7. The power of the test. It is difficult to discuss the power of a non-parametric test, but in the present case it may be worthwhile to give an example or two. The reader will understand that although the test is called non-parametric, its power does depend on the distribution function.

In the case of k samples there are two extremes which might be considered for any particular form of distribution function. In Case 1, we suppose that when the alternative hypothesis is true, $k - 1$ of the populations are identical with distribution function $f(x)$, while the remaining distribution function is $f(x - a)$, $a > 0$. Case 1 may be regarded as a *lower bound* to the power of the test because for any fixed distance a between the location parameters of the rightmost population and the next rightmost population, Case 1 gives the least chance of detecting the falsity of the null hypothesis.

In Case 2, we suppose that the rightmost population is $f(x - a)$, $a > 0$ as before, that the next rightmost population is $f(x)$, and that the other $k - 2$ populations have slipped so far to the left that they make no contribution to problem of the power. This is an optimistic approach to the power because it gives an *upper bound* to the power. When $k = 2$, Case 1 and Case 2 are identical, and the power is exactly the power of the test for the particular distribution function under consideration.

Case 3 which we shall not consider deals with the situation where there is more

than one rightmost population, but the null hypothesis is false. It is connected with the fourth kind of error mentioned at the end of section 6.

Table II gives the upper and lower bound of the power of the test for $k = 3$, $r = 3, n = 3$, when the distribution is uniform and of length unity. The parameter a is the distance between the location parameter of the rightmost distribution and that of the next rightmost distribution.

In Table III we give some points on the upper and lower bounds of the power of the test for the normal distribution with unit standard deviation. The parameter a is the distance between the mean of the rightmost normal distribution and the next rightmost, measured in standard deviations Again we use the case $k = 3$, $r = 3$, $n = 3$.

TABLE II

Power p of the test for the uniform distribution when $k = 3$, $r = 3$, $n = 3$. The distance between the midpoints of the two rightmost distributions is a

a	0	.1	.3	.5	.7	.9	1.00
Upper bound p_u	.05	.09	.23	.46	.73	.96	1 00
Lower bound p_l	.01	.03	.11	.29	.59	.93	1.00

TABLE III

Power p of the test for the unit normal when $k = 3$, $r = 3$, $n = 3$. The distance between the means of the two rightmost distributions, measured in standard deviations, is a

a	0	5	1 0	1.5	2.0	2.5	3.0
Upper bound p_u	.05	.13	.26	.42	.58	.71	.87
Lower bound p_l	.01	.04	.14	.27	.43	.60	.80

The power of the test has been defined as the probability of correctly rejecting the null hypothesis and finding the sample from the rightmost population to be the extreme one. This raises a question about the meaning of the entries in Tables II and III under $a = 0$. When $a = 0$ there is no way to reject the null hypothesis correctly. The probabilities given are the probabilities that a randomly chosen sample will force a rejection of the null hypothesis. They represent the limit of the power function as a tends to zero. If we think of ear-marking the sample from the rightmost population and of computing the probability repeatedly that that sample will have three observations larger than all the observations in the other sample, and then we let a tend to zero, this is the result we get. These values are not the significance levels The significance level is .036.

8. Discussion. The reader may rightly feel that the solution here presented to the problem of the greatest one depends on a trick. That is, it depends intimately on the choice of the null hypothesis. Furthermore the reader may feel that the choice of $a_1 = a_2 = \cdots = a_k$ is neither an interesting null hypothesis nor one which is likely to arise in a practical situation. The author has no quarrel with this attitude. This means that there are many other approaches to this problem which are worth trying. The equal-location-parameter case is one which yields easily to non-parametric methods.

It will be noted that a useful technique has been indicated which allows one to examine the data before making the significance test. In general one may wish to set up a test function, decide which of several samples provides the extreme value of the function, and then test significance given that we have chosen that sample which maximizes the function among the k samples under consideration

9. Conclusion. There is a large class of problems grouped around "the problem of the greatest one". First it would be useful to have a more powerful test than the one here proposed Second, there is the problem of deciding on the basis of samples whether we have successfully predicted the order of the location parameters of several populations Third, there is the general problem of what alternatives, what null hypotheses, and what test functions to use in treating samples from more than two populations. It is to be hoped that more material on these problems will appear, because answers to these questions are urgently needed in practical problems.

ON THE UNIQUENESS OF SIMILAR REGIONS

By Paul G. Hoel

University of California at Los Angeles

1. Summary. Conditions are determined for insuring that Neyman's method of constructing similar regions by means of sufficient statistics will yield all such regions when such statistics exist.

2. Introduction. In designing tests of composite hypotheses, one encounters the problem of how to construct similar regions and whether the construction process yields all possible similar regions. Neyman has derived methods for obtaining similar regions when the basic distribution function satisfies certain partial differential equations [1] and also when a sufficient set of statistics exists for the unknown parameters [2]. In the former case, the construction process gave all such regions; however the question of whether certain subregions were independent of the parameters was left unanswered. In the latter case, the independence was obvious, but the question of uniqueness was not considered. In obtaining sufficient conditions for the existence of a type B region, Scheffé [3] employed Neyman's differential equations assumptions and methods and demonstrated that the subregions were independent of the parameters.

The method of constructing similar regions by means of sufficient statistics is much simpler to demonstrate than is the method based on differential equations. It also has the advantage that the independence of the subregions requires no proof. It possesses the disadvantage that the question of uniqueness is not answered. This question can be answered by showing that the assumption of a sufficient set of statistics includes the differential equations assumption and then employing methods based on the latter assumption. Such a procedure would deprive the sufficiency method of its simplicity, consequently a relatively simple direct proof of uniqueness has been constructed. The method of proof also shows the equivalence of the two methods of constructing similar regions.

3. Sufficient conditions for uniqueness. Consider a distribution function, $f(x|\theta_1, \cdots, \theta_\nu)$, of the variable x that depends upon the ν parameters $\theta_1, \cdots, \theta_\nu$. Let $x_1, x_2, \cdots, x_n$ denote a random sample from this distribution and let $f(x_1, \cdots, x_n|\theta_1, \cdots, \theta_\nu)$ denote the distribution function of such a sample. It will be assumed that $n > \nu$.

Suppose there exists a sufficient set of statistics $T_1(x_1, \cdots, x_n), \cdots, T_\nu(x_1, \cdots, x_n)$ with respect to the parameters $\theta_1, \cdots, \theta_\nu$. Koopman [4] has shown that if the T's are continuous and if $f(x|\theta_1, \cdots, \theta_\nu)$ is analytic, then $f(x|\theta_1, \cdots, \theta_\nu)$ must be a function of the form

$$f(x \mid \theta_1; \cdots; \theta_\nu) = \exp\left[\sum_1^\mu \Theta_k X_k + \Theta + X\right], \tag{1}$$

where the Θ_k and Θ are single-valued analytic functions of the θ's only, and the X_k and X are single-valued analytic functions of x only. He has also shown that if μ assumes its smallest possible value, then

$$\sum_{i=1}^{n} X_k(x_i) = V_k(T_1, \cdots, T_\nu), \tag{2}$$

where the V's are single-valued functions of the T's. If the preceding conditions are satisfied, it follows from (1) and (2) that

$$f(x_1, \cdots, x_n \mid \theta_1, \cdots, \theta_\nu) = \exp\left[\sum_{1}^{\mu} \Theta_k V_k + n\Theta + \sum_{i=1}^{n} X(x_i)\right] \tag{3}$$

Now it is known [2] that if the T's possess continuous partial derivatives and are such that it is possible to introduce additional functions $T_{\nu+1}, \cdots, T_n$ which will make the transformation

$$\begin{aligned} T_1 &= T_1(x_1, \cdots, x_n) \\ &\vdots \\ T_n &= T_n(x_1, \cdots, x_n) \end{aligned} \tag{4}$$

one-to-one, then $f(x_1, \cdots, x_n|\theta_1, \cdots, \theta_\nu)$ can be written in the form

$$\begin{aligned} &f(x_1, \cdots, x_n|\theta_1, \cdots, \theta_\nu) \\ &= f_1(T_1, \cdots, T_\nu|\theta_1, \cdots, \theta_\nu)f_2(x_1, \cdots, x_n|T_1, \cdots, T_\nu), \end{aligned} \tag{5}$$

where f_1 is the distribution function of the T's and f_2 is the conditional distribution function of the x's for fixed values of the T's. The function f_2 does not depend upon any of the parameters $\theta_1, \cdots, \theta_\nu$.

For the purpose of constructing similar regions, it is desirable to work with f_1. By combining (3) and (5), f_1 may be expressed in the form

$$f_1(T_1, \cdots, T_\nu \mid \theta_1, \cdots, \theta_\nu) = \exp\left[\sum_{1}^{\mu} \Theta_k V_k + n\Theta + H\right], \tag{6}$$

where $H = \Sigma X(x_i) - \log f_2$ can be expressed as a function of $T_1, \cdots, T_\nu$ only, and where it is assumed that $f_2 > 0$.

The method employed by Neyman to obtain a similar region of size α is to build it up as the locus of subregions of size α on the "surfaces" obtained by giving the T's constant values. Since the size of such a subregion is obtained by integrating f_2 over the subregion, it will depend only upon the T's; consequently a subregion can be selected that will be of size α for every set of values of the T's.

Now consider the construction of a similar region of size α by building up the region as the locus of subregions of varying size rather than of constant size on the surfaces that are obtained by giving the T's constant values. Let w_1 and w_2 be two regions of size α and let $\alpha_1(T_1, \cdots, T_\nu)$ and $\alpha_2(T_1, \cdots, T_\nu)$ denote the

sizes of the surface subregions. It will be assumed that the regions under consideration are such that α_1 and α_2 are obtainable from integrating f_2 over the subregion common to w_1 and w_2 respectively and the surface determined by fixing the values of the T's. The problem then is to determine whether two different functions, α_1 and α_2, can yield similar regions of size α.

Since a critical region can be obtained as the locus of subregions, α_1 and α_2 will yield similar regions of size α only if

$$\int \cdots \int \alpha_j(T_1, \cdots, T_\nu) f_1(T_1, \cdots, T_\nu \mid \theta_1, \cdots, \theta_\nu)\, dT_1 \cdots dT_\nu = \alpha \tag{7}$$

$$(j = 1, 2),$$

where the integration extends over the range of values of the T's. By means of (6), condition (7) may be written as

$$\int \cdots \int \alpha_j \exp\left[\sum_1^{\mu} \Theta_k V_k + n\Theta + H\right] dT_1 \cdots dT_\nu = \alpha \quad (j = 1, 2). \tag{8}$$

If $e^{n\theta}$ is factored out, it is clear that condition (8) will hold only if

$$\begin{aligned} \int \cdots \int \alpha_1 \exp\left[\sum_1^{\mu} \Theta_k V_k + H\right] dT_1 \cdots dT_\nu \\ = \int \cdots \int \alpha_2 \exp\left[\sum_1^{\mu} \Theta_k V_k + H\right] dT_1 \cdots dT_\nu \end{aligned} \tag{9}$$

is an identity in the θ's, and hence in the Θ_k for the region in the Θ_k space that corresponds to the region in the parameter space for which the parameters $\theta_1, \cdots, \theta_\nu$ are defined.

Now assume that $\mu = \nu$ and that the transformation

$$\begin{aligned} V_1 &= V_1(T_1, \cdots, T_\nu) \\ &\;\;\vdots \\ V_\nu &= V_\nu(T_1, \cdots, T_\nu) \end{aligned} \tag{10}$$

is one-to-one. From the preceding assumptions that gave rise to (2) and (4), it may be shown that the V's are continuous and possess continuous partial derivatives. In terms of the V's, (9) may therefore be written as

$$\begin{aligned} \int \cdots \int \exp\left[\sum_1^{\nu} \Theta_k V_k\right] K_1\, dV_1 \cdots dV_\nu \\ = \int \cdots \int \exp\left[\sum_1^{\nu} \Theta_k V_k\right] K_2\, dV_1 \cdots dV_\nu, \end{aligned} \tag{11}$$

where $K_i = \alpha_i e^H$ has been expressed in terms of the V's.

Since the parameters will be defined over intervals and Θ_k is an analytic function of those parameters, to every region in the parameter space determined by

intervals of the θ's there will correspond an interval for Θ_k throughout which Θ_k will be defined; consequently (11) will be an identity in the Θ_k for intervals of values. For every point within regions determined by Θ_k intervals, the partial derivatives of the two sides of (11) must therefore be equal, provided the derivatives exist and provided the Θ_k are functionally independent.

If the conditions to be imposed shortly are satisfied, it can easily be shown that it is permissible to differentiate (11) repeatedly under the integral signs with respect to the Θ_k. As a consequence, (11) implies that for all sets of non-negative integers $k_1, \cdots, k_\nu$,

$$\begin{aligned}(12)\qquad &\int \cdots \int V_1^{k_1} \cdots V_\nu^{k_\nu} \exp\left[\sum_1^\nu \Theta_k V_k K_1\right] dV_1 \cdots dV_\nu \\ &\qquad = \int \cdots \int V_1^{k_1} \cdots V_\nu^{k_\nu} \exp\left[\sum_1^\nu \Theta_k V_k K_2\right] dV_1 \cdots dV_\nu\end{aligned}$$

will be an identity in the Θ_k for almost all values of the Θ_k. But (12) is equivalent to requiring that

$$\begin{aligned}(13)\qquad &\int \cdots \int V_1^{k_1} \cdots V_\nu^{k_\nu} g_1(V_1, \cdots, V_\nu)\, dV_1 \cdots dV_\nu \\ &\qquad = \int \cdots \int V_1^{k_1} \cdots V_\nu^{k_\nu} g_2(V_1, \cdots, V_\nu)\, dV_1 \cdots dV_\nu\end{aligned}$$

shall hold for all sets of non-negative integers $k_1, \cdots, k_\nu$, where g_1 and g_2 are the integrands of (11) after they have been divided by the function of the Θ_k obtained from integrating (11). Since g_1 and g_2 will then be non-negative functions of the V's whose integrals over all values of the V's is one, they are distribution functions of the V's. If g_1 and g_2 possess moments of all orders and are such that they are uniquely determined by their moments, then condition (13) implies that

$$(14)\qquad g_1(V_1, \cdots, V_\nu) = g_2(V_1, \cdots, V_\nu).$$

This identity will hold for almost all values of the parameters. If the conditions necessary to justify (14) are satisfied, it therefore follows that

$$\alpha_1(T_1, \cdots, T_\nu) = \alpha_2(T_1, \cdots, T_\nu),$$

and that Neyman's method of constructing similar regions by choosing $\alpha(T_1, \cdots, T_\nu) = \alpha$ yields all possible similar regions of the class of regions being considered.

The conditions that were imposed on $f(x|\theta_1, \cdots, \theta_\nu)$ in order to establish uniqueness may be summarized as follows: The distribution function $f(x|\theta_1, \cdots, \theta_\nu)$ is analytic and possesses a set of sufficient statistics, $T_1, \cdots, T_\nu$, with respect to the parameters $\theta_1, \cdots, \theta_\nu$, that are continuous and possess continuous partial derivatives. There exist one-to-one transformations of the types (4) and (10). The function $ce^{\Sigma\Theta_k V_k + H}$, treated as a distribution function of the V's, possesses moments of all orders and is uniquely determined by its moments.

Finally, the Θ_k are functionally independent with the smallest possible value of μ equal to ν.

If the assumption that the Θ_k are independent is not realized, the distribution function (1) could be expressed in terms of fewer than ν parameters. This is also true if $\mu < \nu$. The two assumptions that $\mu = \nu$ and that the Θ_k are independent will therefore be satisfied if (1) is expressed in terms of the minimum number of parameters. The remaining assumptions can often be checked quite easily whenever a particular distribution function is given.

In deriving tests of hypotheses for certain parameters, the distribution function $f(x|\theta_1, \cdots, \theta_\nu)$ will of course contain those parameters in addition to the parameters $\theta_1, \cdots, \theta_\nu$, but since they will have fixed values, it was not necessary to introduce them into the discussion.

4. Equivalence of methods. Although the equivalence of the two methods of constructing similar regions has been implied in the literature [1], no simple demonstration seems to be available. Such a demonstration is easily given by means of (3). Let

$$\varphi_i = \frac{\partial \log f}{\partial \theta_i},$$

where f is given by (3) with $\mu = \nu$, and let

$$\varphi_{ij} = \frac{\partial \varphi_i}{\partial \theta_j}.$$

Differentiation of (3) yields

$$\begin{aligned} \varphi_i &= \sum_1^\nu \frac{\partial \Theta_k}{\partial \theta_i} V_k + n \frac{\partial \Theta}{\partial \theta_i}, \\ \varphi_{ij} &= \sum_1^\nu \frac{\partial^2 \Theta_k}{\partial \theta_i \partial \theta_k} V_k + n \frac{\partial^2 \Theta}{\partial \theta_i \partial \theta_j}. \end{aligned} \tag{15}$$

The differential equations that are assumed to hold in the other method of construction [1] may be written in the form

$$\varphi_{ij} = A_{ij} + \sum_{r=1}^\nu B_{ijr} \varphi_r, \qquad (i, j = 1, \cdots, \nu), \tag{16}$$

where the A_{ij} and B_{ijr} are functions of the θ's only. Upon substituting the values given by (15), it will be found that (16) will be satisfied if

$$\frac{\partial^2 \Theta_k}{\partial \theta_i \partial \theta_j} = \sum_{r=1}^\nu B_{ijr} \frac{\partial \Theta_k}{\partial \theta_r}, \qquad (k = 1, \cdots, \nu) \tag{17}$$

and

$$n \frac{\partial^2 \Theta}{\partial \theta_i \partial \theta_j} = A_{ij} + n \sum_{r=1}^\nu B_{ijr} \frac{\partial \Theta}{\partial \theta_r}.$$

Since (17) represents a set of ν equations in the B_{ij}'s, whose coefficient matrix is non-singular because of the functional independence of the Θ_k, it follows that sets of A's and B's can be found to satisfy equations (16). This shows that the sufficiency assumption includes the differential equations assumption.

Now the method of constructing similar regions here consists in building them up as the locus of subregions of size α on the surfaces obtained by giving the φ_i constant values. But from (15) it follows that the surface $\varphi_i = c_i (i = 1, \cdots, \nu)$ is equivalent to the surface

$$\sum_1^{\nu} \frac{\partial \Theta_k}{\partial \theta_i} V_k + n \frac{\partial \Theta}{\partial \theta_i} = c_i, \qquad (i = 1, \cdots, \nu)$$

which may be written in the form

$$\sum_1^{\nu} \frac{\partial \Theta_k}{\partial \theta_i} V_k = c_i', \qquad (i = 1, \cdots, \nu), \tag{18}$$

because Θ is a function of the parameters only. Since the coefficient matrix of the V's in (18) is nonsingular, (18) may be solved for the V's, consequently the surface $\varphi_i = c_i$, $(i = 1, \cdots, \nu)$ is equivalent to the surface $V_i = c_i''$, $(i = 1, \cdots, \nu)$. But from the assumption concerning the transformation (10), the surface $V_i = c_i''$, $(i = 1, \cdots, \nu)$ is equivalent to the surface $T_i = c_i'''$, $(i = 1, \cdots, \nu)$. Thus, the two surfaces $\varphi_i = c_i$ $(i = 1, \cdots, \nu)$ and $T_i = c_i'''$, $(i = 1, \cdots, \nu)$ are equivalent and hence the two methods of constructing similar regions are equivalent.

REFERENCES

[1] J. NEYMAN, "On a statistical problem arising in routine analysis and in sampling inspections of mass production," *Annals of Math. Stat*, Vol. 12 (1941), pp. 46–76.

[2] J. NEYMAN, "Outline of a theory of statistical estimation based on the classical theory of probability," *Roy Soc. Phil. Trans.*, Vol 236A (1937), p. 364.

[3] H. SCHEFFÉ, "On the theory of testing composite hypotheses with one constraint," *Annals of Math. Stat.*, Vol. 13 (1942), pp 280–293.

[4] B. O. KOOPMAN, "On distributions admitting a sufficient statistic," *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 399–409.

NOTES

This section is devoted to brief research and expository articles and other short items.

CONVERGENCE OF DISTRIBUTIONS

By Herbert Robbins

University of North Carolina

Let $f_n(x)$ $(n = 0, 1, 2, \cdots)$ be frequency functions

$$f_n(x) \geq 0, \qquad \int_{-\infty}^{\infty} f_n(x)\, dx = 1. \tag{1}$$

There are various ways in which the sequence of distributions corresponding to the $f_n(x)$ $(n = 1, 2, \cdots)$ may be said to converge to the distribution corresponding to $f_0(x)$. The definition customarily adopted in mathematical statistics (see e.g. [1]) is equivalent to the condition

$$\lim_{n\to\infty} \int_{-\infty}^{\xi} f_n(x)\, dx = \int_{-\infty}^{\xi} f_0(x)\, dx \qquad \text{for every } \xi.^1 \tag{a}$$

We shall also consider the two further conditions

$$\lim_{n\to\infty} \int_S f_n(x)\, dx = \int_S f_0(x)\, dx \qquad \text{for every Borel set } S, \tag{b}$$

and

$$\lim_{n\to\infty} \int_S f_n(x)\, dx = \int_S f_0(x)\, dx \qquad \text{uniformly for all Borel sets } S. \tag{c}$$

It is clear that (c) implies (b) and that (b) implies (a). That the converse implications do not hold is shown by the following examples.

Example 1. Let $f_0(x) = 1$ for $0 \leq x \leq 1$ and 0 elsewhere. Choose and fix any $0 < \epsilon < 1$, set $\delta_n = \epsilon/n \cdot 2^n$, and for $n = 1, 2, \cdots$ let $f_n(x) = 1/n \cdot \delta_n$ for $i/n - \delta_n \leq x \leq i/n$ $(i = 1, 2, \cdots, n)$ and 0 elsewhere. If we denote by S_n the set of all x for which $f_n(x) > 0$ it is easy to see that for $n = 1, 2, \cdots$

$$0 \leq \int_{-\infty}^{\xi} f_0(x)\, dx - \int_{-\infty}^{\xi} f_n(x)\, dx < 1/n \qquad \text{for every } \xi, \tag{2}$$

$$\int_{S_n} f_0(x)\, dx = \epsilon/2^n, \qquad \int_{S_n} f_n(x)\, dx = 1. \tag{3}$$

---

[1] From a well kown theorem of Pólya the convergence is then necessarily uniform for all ξ.

Hence for the Borel set $S = \sum_1^\infty S_n$ it follows that

$$\int_S f_0(x)\,dx \leq \sum_1^\infty \int_{S_n} f_0(x)\,dx = \epsilon, \tag{4}$$

$$\int_S f_n(x)\,dx = \int_{S_n} f_n(x)\,dx = 1, \qquad (n = 1, 2, \cdots). \tag{5}$$

From (2) we see that (a) holds (uniformly for all ξ), and from (4) and (5) that (b) fails about as badly as possible.

This construction can be modified to apply to any $f_0(x)$; thus choosing $f_0(x) = (2\pi e^{x^2})^{-1/2}$ we can construct $f_n(x)$ $(n = 1, 2, \cdots)$ and a Borel set S such that

$$\lim_{n\to\infty} \int_{-\infty}^{\xi} f_n(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\xi} e^{-x^2/2}\,dx \qquad \text{uniformly for all } \xi,$$

while

$$\frac{1}{\sqrt{2\pi}} \int_S e^{-x^2/2}\,dx = .01, \qquad \int_S f_n(x)\,dx = 1, \qquad (n = 1, 2, \cdots).$$

It is conceivable that some time a statistician, failing to consider such a possibility, will be led to approximate .01 by 1.

If X_n is a random variable with frequency function $f_n(x)$, if $y = g(x)$ is a Borel function, and if (a) holds, then it follows from Example 1 that the distribution function $H_n(y)$ of $Y_n = g(X_n)$, equal to the integral of $f_n(x)$ over the set S_y of all x such that $g(x) \leq y$, need not converge to the distribution function $H_o(y)$ of $Y_o = g(X_o)$. It is easily seen that this possibility is excluded if, as commonly occurs in applications, $g(x)$ is such that for every y, the intersection of S_y with any finite interval is the sum of a finite number of intervals (e.g., if $g(x) = \sin x$).

EXAMPLE 2. Let $f_0(x)$ be defined as in the previous example, and for $n = 1, 2, \cdots$ let $f_n(x) = 1 + \sin(2\pi n x)$ for $0 \leq x \leq 1$ and 0 elsewhere. By the Riemann-Lebesgue theorem it follows that (b) holds. But let S_n denote the set of all x for which $f_n(x) \geq 1$; then

$$\int_{S_n} f_0(x)\,dx = \tfrac{1}{2}, \qquad \int_{S_n} f_n(x)\,dx = \tfrac{1}{2} + 1/\pi, \qquad (n = 1, 2, \cdots),$$

so that (c) does not hold.

It follows from these examples that (a), (b), and (c) are successively stronger definitions of convergence. We shall now give some definitions equivalent to (b) and (c).

First we recall that the non-negative, completely additive, and absolutely continuous set functions

$$P_n(S) = \int_S f_n(x)\,dx, \qquad (n = 1, 2, \cdots), \tag{6}$$

are said to be uniformly absolutely continuous if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for any S and any $n = 1, 2, \cdots$,

$$m(S) < \delta \text{ implies } P_n(S) < \epsilon. \tag{7}$$

We shall denote the condition that the $P_n(S)$ be uniformly absolutely continuous by (u.a.c.), and we shall now prove that (b) is equivalent to

(b′) (a) and (u.a.c.).

PROOF. (A) Suppose (b) holds. It is clear that (a) holds, and we shall show by contradiction that (u.a.c.) holds also. For if not then there would exist an $\epsilon > 0$ such that for any $\eta > 0$ we could find a set S and an integer n such that

$$m(S) < \eta, \qquad P_n(S) \geq \epsilon. \tag{8}$$

Moreover, since the set function

$$P_0(S) = \int_S f_0(x)\, dx$$

is absolutely continuous, there exists a $\delta > 0$ such that

$$m(S) < \delta \quad \text{implies} \quad P_0(S) < \epsilon/2. \tag{9}$$

Now by (8) there exists an S_1 with $m(S_1) < \delta/2$ and a k_1 such that $P_{k_1}(S_1) \geq \epsilon$. Next, there exists an S_2 with $m(S_2) < \delta/2^2$ and a k_2 such that $P_{k_2}(S_2) \geq \epsilon$, and it is easy to see that we may assume that $k_2 > k_1$. Proceeding in this way we find a sequence of integers $k_1 < k_2 < \cdots$ and of sets $S_1, S_2, \cdots$ such that

$$m(S_n) < \delta/2^n, \qquad P_{k_n}(S_n) \geq \epsilon, \qquad (n = 1, 2, \cdots). \tag{10}$$

Let $S = \sum_1^\infty S_n$; then by (10), $m(S) \leq \sum_1^\infty m(S_n) < \delta$, so that by (9),

$$P_0(S) < \epsilon/2. \tag{11}$$

But by (10),

$$P_{k_n}(S) \geq P_{k_n}(S_n) \geq \epsilon, \qquad (n = 1, 2, \cdots). \tag{12}$$

From (11) and (12) we conclude that (b) does not hold, which is a contradiction. Hence (b) implies (b′).

(B) Suppose (b′) holds. We shall show first that (b) holds for any set S_1 of finite measure. Choose any $\epsilon > 0$; then from (u.a.c.) it follows that there exists a $\delta > 0$ such that

$$m(S) < \delta \text{ implies } P_n(S) < \epsilon/8 \qquad (n = 0, 1, 2, \cdots). \tag{13}$$

It is known from the theory of measure that corresponding to S_1 and to δ we can find a set S_2 which is the sum of a finite number of disjoint intervals, such that

$$m((S_1 - S_2) + (S_2 - S_1)) < \delta. \tag{14}$$

From (13), (14), and the relations

$$(15) \quad P_n(S_1) = P_n(S_2) + P_n(S_1 - S_2) - P_n(S_2 - S_1), \qquad (n = 0, 1, 2, \cdots),$$

it follows that

$$(16) \quad \begin{aligned} | P_0(S_1) - P_n(S_1) | \leq | P_0(S_2) - P_n(S_2) | + P_n(S_1 - S_2) + P_n(S_2 - S_1) \\ + P_0(S_1 - S_2) + P_0(S_2 - S_1) \leq | P_0(S_2) - P_n(S_2) | + \epsilon/2, \end{aligned}$$

and from (a) that for large enough n,

$$(17) \qquad | P_0(S_2) - P_n(S_2) | < \epsilon/2$$

Thus from (16) and (17) it follows that for large enough n,

$$| P_0(S_1) - P_n(S_1) | < \epsilon,$$

which proves (b) for the case $m(S) < \infty$.

Now given any $\epsilon > 0$ choose α, β so that, setting $A = \{\alpha \leq x \leq \beta\}$, we have

$$(19) \qquad P_0(A) > 1 - \epsilon/4.$$

Then it follows from (a) that for large enough n,

$$(20) \qquad P_n(A) > 1 - \epsilon/2.$$

Then for any Borel set S we have for large enough n,

$$P_n(S) - P_0(S) = P_n(SA) + P_n(S - A) - P(SA) - P(S - A),$$

$$\begin{aligned} | P_n(S) - P_0(S) | &\leq | P_n(SA) - P_0(SA) | + P_n(S - A) + P_0(S - A) \\ &\leq | P_n(SA) - P_0(SA) | + \epsilon/2 + \epsilon/4. \end{aligned}$$

But by the previous case, since $m(SA) < \infty$, for large enough n we shall have $| P_n(SA) - P_0(SA) | < \epsilon/4$. Hence for large enough n,

$$| P_n(S) - P_0(S) | < \epsilon,$$

so that (b) holds in this case also. This completes the proof.

We shall say that $\lim_{n \to \infty} f_n(x) = f_0(x)$ in measure if for every $\epsilon > 0$ and for every set A such that $m(A) < \infty$, the measure of the set of all x in A for which $|f_n(x) - f_0(x)| > \epsilon$, tends to 0 as n increases. (For a space of finite measure this reduces to the usual definition.) We now observe that (c) is equivalent to

$$(c') \qquad \lim_{n \to \infty} f_n(x) = f_0(x) \qquad \text{in measure.}$$

In fact, it is easy to show that (c) is equivalent to convergence in the mean of order 1,

$$(c'') \qquad \lim_{n \to \infty} \int_{-\infty}^{\infty} | f_n(x) - f_0(x) | \, dx = 0,$$

which implies (c′), and a theorem of Scheffé [2] states that (c′) implies (c).[2] Finally, it is not hard to show that the condition

$$\text{(d)} \qquad \lim_{n\to\infty} f_n(x) = f_0(x) \qquad \text{almost everywhere}$$

implies (c′) but not conversely.

Summing up, we arrive at the following complete set of implication relations among the various modes of convergence which we have considered:

$$\text{(20)} \qquad (d) \rightarrow (c'') \rightleftarrows (c') \rightleftarrows (c) \rightarrow (b') \rightleftarrows (b) \rightarrow (a).$$

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, 1946, pp. 58–60.
[2] H. Scheffé, "A useful convergence theorem for probability distributions," *Annals of Math Stat.*, Vol. 18 (1947), pp. 434–438.
[3] E. J. McShane, *Integration*, Princeton Univ. Press, 1944, p. 168.

ON RANDOM VARIABLES WITH COMPARABLE PEAKEDNESS

By Z. W. Birnbaum

University of Washington

The quality of a distribution usually referred to as its peakedness has often been measured by the fourth moment of the distribution. It is known, however, that there is no definite connection between the value of the fourth moment and what one may intuitively consider as the amount of peakedness of a distribution.[1] In the present paper a definition of relative peakedness is proposed and it is shown that this concept has properties which may make it practically applicable.

Definition. *Let Y and Z be real random variables and Y_1 and Z_1 real constants. We shall say that Y is more peaked about Y_1 than Z about Z_1 if the inequality*

$$P(|Y - Y_1| \geqq T) \leqq P(|Z - Z_1| \geqq T)$$

is true for all $T \geqq 0$.

If, for example, Y and Z are normal random variables with expectations Y_1 and Z_1 and standard deviations σ_y and σ_z, and if $\sigma_y < \sigma_z$, then Y is more peaked about Y_1 than Z about Z_1. Similarly, if Y is a random variable such that $P(Y < a) = P(Y > b) = 0$ for $a < b$, and if Z is the discrete random variable with $P(Z = a) = P(Z = b) = \frac{1}{2}$, then Y is more peaked about $\frac{1}{2}(a + b)$ than Z about the same point.

---

[2] Scheffé actually proves that (d) implies (c), but the Lebesgue convergence theorem on which his proof is based holds for convergence in measure (see e.g. [3]).

[1] I. Kaplansky, "A common error concerning kurtosis," *Am. Stat. Assn. Jour.*, Vol. 40 (1945), p. 259.

LEMMA *Let Y_1, Y_2, Z_1, Z_2 be continuous random variables*[2] *with the probability densities $\varphi_1(Y_1)$, $\varphi_2(Y_2)$, $f_1(Z_1)$, $f_2(Z_2)$ such that*

1^0. *Y_1 and Y_2 are independent, Z_1 and Z_2 are independent,*

2^0 *$\varphi_i(Y_i) = \varphi_i(-Y_i)$ for all Y_i, $f_i(Z_i) = f_i(-Z_i)$ for all Z_i, $(i = 1, 2)$,*

3^0. *$\varphi_2(Y_2)$ and $f_1(Z_1)$ are not-increasing functions for positive values of the variables,* and

4^0. *Y_i is more peaked about 0 than Z_i, for $i = 1, 2$.*

Let $Y = Y_1 + Y_2$ and $Z = Z_1 + Z_2$. Under these assumptions Y is more peaked about 0 than Z.

PROOF: Let $\Phi_i(y) = P(Y_i \leqq y)$, $F_i(z) = P(Z_i \leqq z)$, for $i = 1, 2$, be the cumulative probability functions. For any random variables Y_1, Y_2, Z_1, Z_2 (not necessarily continuous) which fulfil assumption 1^0 we have, for any T, the relationships

$$
\begin{aligned}
P(Y \leqq T) - P(Z \leqq T) &= \int_{-\infty}^{\infty} [\Phi_1(T - s)d\Phi_2(s) - F_1(T - s)dF_2(s)] \\
&= \int_{-\infty}^{\infty} [\Phi_1(T - s) - F_1(T - s)]d\Phi_2(s) \\
&\qquad + \int_{-\infty}^{\infty} F_1(T - s)[d\Phi_2(s) - dF_2(s)] \\
&= \int_{-\infty}^{\infty} [\Phi_1(T - s) - F_1(T - s)]d\Phi_2(s) \\
&\qquad - \int_{-\infty}^{\infty} [\Phi_2(s) - F_2(s)]dF_1(T - s) \\
&= \int_{-\infty}^{\infty} [\Phi_1(T - s) - F_1(T - s)]d\Phi_2(s) \\
&\qquad + \int_{-\infty}^{\infty} [\Phi_2(T - s) - F_2(T - s)]dF_1(s) \\
&= I_1(T) + I_2(T),
\end{aligned}
$$

where

$$
\begin{aligned}
I_1(T) &= \int_{-\infty}^{\infty} [\Phi_1(T - s) - F_1(T - s)]d\Phi_2(s) \\
&= \int_{-\infty}^{\infty} [\Phi_1(-s) - F_1(-s)]d\Phi_2(T + s) \\
&= \int_{-\infty}^{0} + \int_{0}^{\infty} \\
&= \int_{0}^{\infty} \{[F_1(s) - \Phi_1(s)]d\Phi_2(T - s) \\
&\qquad + [\Phi_1(-s) - F_1(-s)]d\Phi_2(T + s)\},
\end{aligned}
$$

etc.

---

[2] As defined e.g. in H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 169.

If the random variables have distributions symmetrical about zero (assumption 2^0) this is equal to

$$\int_0^{+\infty} \{[P(Z_1 \leqq s) - P(Y_1 \leqq s)]dP(Y_2 \leqq T - s) + [P(Y_1 \leqq -s) - P(Z_1 \leqq -s)]dP(Y_2 \leqq T + s)\}$$

$$= \int_0^{+\infty} \{[1 - P(Z_1 > s) - 1 + P(Y_1 > s)]dP(Y_2 \leqq T - s) + [P(Y_1 \geqq s) - P(Z_1 \geqq s)]dP(Y_2 \leqq T + s)\}$$

$$= \int_0^{+\infty} \{[P(Y_1 \geqq s) - P(Z_1 \geqq s)]d[P(Y_2 \leqq T + s) + P(Y_2 \leqq T - s)] - [P(Y_1 = s) - P(Z_1 = s)]dP(Y_2 \leqq T - s)\},$$

and we obtain

$$I_1(T) = \int_0^{+\infty} [P(Y_1 \geqq s) - P(Z_1 \geqq s)]d[P(Y_2 \leqq T + s) + P(Y_2 \leqq T - s)] - \int_0^{+\infty} [P(Y_1 = s) - P(Z_1 = s)]dP(Y_2 \leqq T - s). \tag{1.1}$$

By an analogous argument one derives the equality

$$I_2(T) = \int_0^{+\infty} [P(Y_2 \geqq s) - P(Z_2 \geqq s)]dP(Z_1 \leqq T + s) + P(Z_1 \leqq T - s)] - \int_0^{+\infty} [P(Y_2 = s) - P(Z_2 = s)]dP(Z_1 \leqq T - s). \tag{1.2}$$

Making use of the assumption that Y_1, Y_2, Z_1, Z_2, are continuous random variables, we conclude that the second integrals in (1.1) and (1.2) are zero, and we may write

$$I_1(T) = \int_0^{+\infty} [P(Y_1 \geqq s) - P(Z_1 \geqq s)][\varphi_2(T + s) - \varphi_2(T - s)]ds, \tag{2.1}$$

$$I_2(T) = \int_0^{+\infty} [P(Y_2 \geqq s) - P(Z_2 \geqq s)][f_1(T + s) - f_1(T - s)]ds. \tag{2.2}$$

For $T \geqq 0$ we have, making use of assumption 3^0,

$$\varphi_2(T + s) - \varphi_2(T - s) \leqq 0 \text{ if } 0 \leqq s \leqq T$$

$$\varphi_2(T + s) - \varphi_2(T - s) = \varphi_2(s + T) - \varphi_2(s - T) \leqq 0 \text{ if } 0 \leqq T \leqq s,$$

and similarly

$$f_1(T + s) - f_1(T - s) \leqq 0 \text{ for all } T \geqq 0 \text{ and } s \geqq 0.$$

Since according to assumption 4^0 we also have

$$P(Y_1 \geqq s) - P(Z_1 \geqq s) \leqq 0$$

$$P(Y_2 \geqq s) - P(Z_2 \geqq s) \leqq 0 \text{ for } s \geqq 0,$$

both integrands in (2.1) and (2.2) are non-negative for all values of s, and we conclude

$$P(Y \leqq T) - P(Z \leqq T) = I_1(T) + I_2(T) \geqq 0,$$

and hence

$$P(Y \geqq T) - P(Z \geqq T) \leqq 0 \text{ for } T \geqq 0 \tag{3.1}$$

From assumption 2^0 one easily sees that Y and Z have symmetrical probability distributions. This together with (3.1) leads to

$$P(Y \geqq T) - P(Z \geqq T) = P(Y \leqq -T) - P(Z \leqq -T) \leqq 0,$$

and thus to

$$P(|Y| \geqq T) - P(|Z| \geqq T) \leqq 0 \text{ for } T \geqq 0.$$

As can be seen from (1.1) and (1 2), the assumptions of the Lemma, in particular the assumption that all variables are continuous and the assumption 3^0, are rather special sufficient conditions for Y being more peaked about 0 than Z.

THEOREM 1. *Let Y and Z be continuous random variables with probability densities $\varphi(Y)$ and $f(Z)$ such that*

1^0. *$\varphi(-Y) = \varphi(Y)$ for all Y, $f(-Z) = f(Z)$ for all Z,*

2^0. *$\varphi(Y)$ and $f(Z)$ are not-increasing functions for positive values of the variables,*

3^0. *Y is more peaked about 0 than Z.*

Let $Y_1, Y_2, \cdots, Y_n$ and $Z_1, Z_2, \cdots, Z_n$ be random samples of Y and Z, respectively, and $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$, $\bar{Z}_n = \frac{1}{n}\sum_{j=1}^{n} Z_j$.—Then $\bar{Y}_n$ is more peaked about 0 than $\bar{Z}_n$.

PROOF From the preceding Lemma one concludes by simple induction that $Y' = Y_1 + Y_2 + \cdots + Y_n$ as well as $Z' = Z_1 + Z_2 + \cdots + Z_n$ are continuous random variables with distributions symmetrical about zero and probability densities not-increasing for positive values of the variables, such that Y' is more peaked about 0 than Z'. From this the theorem follows immediately.

The conjecture that assumption 2^0 of Theorem 1 might be superfluous is incorrect as may be seen from the following example:

Let Y be any continuous random variable with a distribution symmetrical about zero and such that $P(|Y| > a) = 0$ for some $a > 0$. Let Z be the discrete random variable with $P(Z = -a) = P(Z = a) = \frac{1}{2}$. We have for $0 \leqq T \leqq a$

$$P(|Y| \geqq T) \leqq 1 = P(|Z| \geqq T),$$

hence Y is more peaked about 0 than Z. If Y_1, Y_2 and Z_1, Z_2 are random samples of size 2, we have

$$P(\bar{Z}_2 = -a) = P(\bar{Z}_2 = a) = \tfrac{1}{4}, \qquad P(\bar{Z}_2 = 0) = \tfrac{1}{2},$$

and thus

$$P(|\bar{Z}_2| \geqq T) = \tfrac{1}{2} \text{ for } 0 < T \leqq a$$

The random variable $\bar{Y}_2$ is continuous, with a distribution symmetrical about zero, such that $P(|\bar{Y}_2| \leqq a) = 1$. There exists, therefore, a T_1 such that $0 < T_1 \leqq a$ and that $P(|\bar{Y}_2| \geqq T_1) = \frac{3}{4}$ It follows that

$$P(|\bar{Y}_2| \geqq T_1) = \tfrac{3}{4} > \tfrac{1}{2} = P(|\bar{Z}_2| \geqq T_1),$$

hence $\bar{Y}_2$ is not more peaked about zero than $\bar{Z}_2$. The random variable Z is discrete, but it can be approximated by a continuous random variable with a U-shaped probability density, so that all the probabilities will be modified only very slightly and $\bar{Y}_2$ still will not be more peaked than $\bar{Z}_2$. Nothing will change in this example if one assumes that Y fulfils condition 2^0 of Theorem 1.

THEOREM 2 *Let Y be a continuous random variable such that*

1^0. $\varphi(-Y) = \varphi(Y)$ *for all* Y,

2^0. $\varphi(Y)$ *is a not-increasing function for* $Y > 0$,

3^0. $P(|Y| > a) = 0$ *for some* $a > 0$.

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample of size n and $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$. Then, for any $y \geqq 0$, we have

$$P(|\bar{Y}_n| \geqq y) \leqq \Psi_n\left(\frac{y}{a}\right), \tag{4.1}$$

where

$$\Psi_n(t) = \frac{2}{n} \sum_{(n/2)(t+1) < k \leqq n} (-1)^k \binom{n}{k} \left[\frac{n}{2}(t+1) - k\right]^n. \tag{4.2}$$

PROOF. Let Z be the random variable with uniform distribution in the interval $-1 \leqq Z \leqq 1$. If $Z_1, Z_2, \cdots, Z_n$ is a random sample, then $Z' = Z_1 + Z_2 + \cdots + Z_n$ has the cumulative probability function[3]

$$P(Z' \leqq z) \begin{cases} = 0, & z < -n, \\ = \dfrac{1}{n!} \displaystyle\sum_{i \leqq (z+n)/2} (-1)^i \binom{n}{i} \left(\frac{z+n}{2} - i\right)^n, & -n \leqq z \leqq n, \\ = 1, & z > n, \end{cases}$$

and $\bar{Z}_n = \dfrac{z'}{n}$ has the cumulative probability function

$$P(\bar{Z}_n \leqq \zeta) \begin{cases} = 0, & \zeta < -1, \\ = \dfrac{1}{n!} \displaystyle\sum_{i \leqq (n/2)(\zeta+1)} (-1)^i \binom{n}{i} \left[\frac{n}{2}(\zeta + 1) - i\right]^n, & -1 \leqq \zeta \leqq 1, \\ = 1, & \zeta > 1. \end{cases}$$

[3] This expression is due to Laplace. For derivation and discussion, see: J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, 1937, p. 279, and Cramér, op. cit., p. 245.

Thus,

$$P(|\bar{Z}_n| \geqq t) = 2[1 - P(\bar{Z}_n \leqq t)]$$
$$= 2\left\{1 - \frac{1}{n!} \sum_{i \leqq (n/2)(t+1)} (-1)^i \binom{n}{i} \left[\frac{n}{2}(t+1) - i\right]^n\right\},$$

and in view of the identity

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} (u - k)^n = n!$$

this becomes

$$P(|\bar{Z}_n| \geqq t) = \frac{2}{n!} \sum_{(n/2)(t+1) < k \leqq n} (-1)^k \binom{n}{k} \left[\frac{n}{2}(t+1) - k\right]^n = \Psi_n(t)$$

for $0 \leqq t \leqq 1$. The random variable $\frac{Y}{a}$ is obviously more peaked about zero than Z. Since $\frac{Y}{a}$ and Z fulfil the assumptions of Theorem 1, it follows that $\frac{\bar{Y}_n}{a}$ is more peaked about zero than $\bar{Z}_n$, that is

$$P\left(\left|\frac{\bar{Y}_n}{a}\right| \geqq t\right) \leqq P(|\bar{Z}_n| \geqq t) = \Psi_n(t) \quad \text{for} \quad t \geqq 0.$$

Setting $at = y$, one obtains (4.1).

For $n \to \infty$ the function $\Psi_n(t)$ approaches asymptotically the probability $P(|X| \geqq t\sqrt{3n})$ for the normalized normal random variable X.[4] For $n = 8$ one obtains the following values which indicate a good approximation:

t	.3998	.5254	.6711
$P(\lvert X \rvert \geqq t\sqrt{24})$	.05	.01	.001
$\Psi_8(t)$	.049	.0092	.0005.

For smaller values of n, $\Psi_n(t)$ can be easily computed.

A METHOD FOR OBTAINING RANDOM NUMBERS

By H. Burke Horton

Interstate Commerce Commission

The need for large quantities of random numbers to be used in sample design, subsampling, and other statistical problems is well known. Tippett's [1] numbers have been widely used for these purposes, despite criticism directed at their lack of randomness The following procedure may be of interest to those

[4] Cramér, op. cit., p. 245

who wish to develop their own random series. The method described below will ultimately be used to record extensive tables of random numbers for general use.

Current methods of producing random numbers usually depend upon single operations of mechanical or electronic devices. These may be described as "single-stage" random number processes. The numerical results are biased to the same extent as the devices from which they are taken.

At this point it is desirable to describe a process which may be called "compound" randomization. Assume two roulette wheels arranged in series so that the first controls the arrangement of symbols on the second wheel, while a turn of the second wheel determines which of its positions is to be observed. If the decimal system is used, the first wheel would have 10! "equally likely" positions, and the second would have 10 "equally likely" positions. If three such wheels were to be chained, the first would require (10!)! positions, the second 10! positions, and the third 10 positions. In general, if n wheels were to be chained, the first would require $10(!)^{n-1}$ "equally likely" positions. It is not practical to design such a machine.[1]

One method of surmounting these difficulties is to shift to the binary system in order to take advantage of the fact that $2! = 2$; or, in general, $2(!)^n = 2$. This property makes feasible the chaining of any number of machines in series; and, furthermore, the machines can be of the same design. If desired, the results taken from a single machine may be chained. Another important feature is the ease of handling binary chains by electronic systems.

The words "equally likely" have been placed in quotation marks thus far to indicate that the probabilities are as nearly equal as manufacturing precision permits. Any simple single-stage device will have some bias, and it is this very lack of true equality that the chaining process is designed to meet. For convenience we may take as our binary symbols $+1$ and -1 rather than the customary 1 and 0. We adhere to the usual rules regarding the sign of a product.

Let p_i be the probability of obtaining $+1$ in the i^{th} trial (or in the i^{th} machine of a chain of machines). $0 \leq p_i \leq 1$. $q_i = 1 - p_i$ represents the probability of obtaining -1 in the i^{th} trial.

Let P_i be the probability of obtaining $+1$ as the *product* of i trials. $Q_i = 1 - P_i$ is the probability of obtaining -1 as the *product* of i trials. The following relationships can be set down immediately:

$$
\begin{array}{ll}
P_1 = p_1 & Q_1 = q_1 \\
P_2 = P_1 \cdot p_2 + Q_1 \cdot q_2 & Q_2 = P_1 \cdot q_2 + Q_1 \cdot p_2 \\
P_3 = P_2 \cdot p_3 + Q_2 \cdot q_3 & Q_3 = P_2 \cdot q_3 + Q_2 \cdot p_3 \\
\cdot & \cdot \\
\cdot & \cdot \\
\cdot & \cdot \\
P_i = P_{i-1} \cdot p_i + Q_{i-1} \cdot q_i & Q_i = P_{i-1} \cdot q_i + Q_{i-1} \cdot p_i
\end{array}
$$

[1] It has been pointed out by Dr. George W. Brown that a practical solution is possible using any number base, n, by addition of random digits $(0, 1, 2, \cdots n - 1)$ modulo n.

We may calculate the bias, $P_k - \frac{1}{2}$, for a chain of k trials:

$$P_k - \tfrac{1}{2} = \tfrac{1}{2}(P_k - Q_k)$$
$$= \tfrac{1}{2}(P_{k-1}\cdot p_k + Q_{k-1}\cdot q_k - P_{k-1}\, q_k - Q_{k-1}\cdot p_k)$$

Factoring, we have

$$P_k - \tfrac{1}{2} = \tfrac{1}{2}(P_{k-1} - Q_{k-1})(p_k - q_k)$$

Substituting for $P_{k-1} - Q_{k-1}$ and factoring again,

$$P_k - \tfrac{1}{2} = \tfrac{1}{2}(P_{k-2} - Q_{k-2})(p_{k-1} - q_{k-1})(p_k - q_k)$$

Continuing the process of substituting and factoring, we obtain

$$P_k - \tfrac{1}{2} = \tfrac{1}{2}(p_1 - q_1)(p_2 - q_2)\cdots(p_k - q_k)$$

(1)
$$P_k - \tfrac{1}{2} = \frac{1}{2}\prod_{i=1}^{k}(p_i - q_i) = \frac{1}{2}\prod_{i=1}^{k}(2p_i - 1).$$

We may write the general formula for P_k:

(2)
$$P_k = \frac{1}{2}\left[1 + \prod_{i=1}^{k}(2p_i - 1)\right].$$

In the special case where all the p_i are equal to a constant, p,

(3)
$$P_k = \tfrac{1}{2}[1 + (2p - 1)^k].$$

This can also be derived directly by expansion of $(p - q)^k$.

If any machine, r, in the chain has no bias ($p_r = \frac{1}{2}$, exactly), the chain itself has no bias, since $2p_r - 1 = 0$. Note also that if for all i, $0 < p_i < 1$, the bias of the complete chain is less than the bias of any component (single or multiple) taken from the chain, because $|(2p_i - 1)| < 1$. Or stated another way, the results taken from any machine, no matter how nearly perfect, can be improved by chaining with another machine, no matter how biased the latter. Even in the limiting case, $p = 1$ (or 0), the magnitude of the bias remains unchanged; in all other cases it is reduced. The bias of final results can be made as small as desired by increasing the length of the chain. Compound randomization can be regarded as an attrition process which may be used to reduce final bias below any preassigned quantity. If the observations taken from two machines in the chain should be perfectly correlated, the only effect is to shorten the chain by two.

In shifting from the binary system to the decimal system, symbol bias will be introduced. In general, symbol bias will be introduced in passing from a given positional system to any other positional system, unless one of the number bases is a rational power of the other.

To illustrate, let us assume that we have a random binary series and wish to obtain a random one-digit decimal series. It will be necessary to tabulate the binary series in blocks of four symbols. The quantities will range from 0000 (binary) to 1111 (binary), or from 00 (decimal) to 15 (decimal), with equal

probabilities. There would be no predominance of either ones or zeros in the overall binary tabulation, as illustrated in the table below.

	Binary System	Decimal System
	0000 0001 0010 0011 0100 0101 0110 0111 1000 1001	0 1 2 3 4 5 6 7 8 9
Tabulation to this point	25 zeros 15 ones	One of each symbol
	1010 1011 1100 1101 1110 1111	10 11 12 13 14 15
Overall tabulation	32 zeros 32 ones	(Right digit only) 0–5, 2 each 6–9, 1 each

However, if we look at the right digit of the decimal tabulation, it is clear that the symbols 0 to 5, inclusive, will occur twice as often as the symbols 6 to 9, inclusive. The easiest way of correcting for this bias is simply to reject all two-digit decimal numbers which occur, thereby giving equal probabilities to the ten decimal symbols. The rejection could be accomplished most easily by electronic devices operating on the binary numbers. All numbers greater than 1001 (binary) would be excluded through the operation of a simple four-stage electronic counter.

This simple illustration also demonstrates the inefficiency of converting random four-digit binary numbers to random one-digit decimal numbers. 37.5% of the data are lost in the process of removing bias. A more efficient procedure would be to tabulate the random binary series in blocks of ten digits. The largest number that could occur would be 1 111 111 111 (binary), or 1,023 (deci-

mal). The numbers would have equal probabilities insofar as this is attainable by chaining. To obtain a random three-digit decimal series it would be necessary to reject the numbers above 999 (decimal). This would amount to only 2.34% of the available data. As before, rejection could be accomplished easily in the binary series by use of a ten-stage electronic counter

Several promising devices are being considered for tabulating random numbers in accordance with the principles discussed herein Electronic or electrical systems actuated by cosmic rays seem to be the most desirable. Tabulating equipment may be wired to turn out random numbers, possibly as a by-product of other card runs.

If only a few random numbers are needed, they can be obtained by much simpler methods. For example, a coin may be tossed, letting heads and tails represent $+1$ and -1, respectively The product of k successive tosses would be tabulated as the random binary variable. Products equal to $+1$ and -1 would be coded as 1 and 0, respectively. Blocks of binary symbols would then be converted to the decimal system as described above.

REFERENCE

[1] Tippett, L. H. C., *Random Sampling Numbers*, Tracts for Computers, No. 15, Cambridge University Press, 1927.

NOTE ON THE ERROR IN INTERPOLATION OF A FUNCTION OF TWO INDEPENDENT VARIABLES

By W. M. Kincaid

University of Michigan

Suppose that g is a functon of one real variable x and h is an interpolation function such that $g(x) = h(x)$ for $x = x_1, x_2, \cdots, x_n$. Let $f(x) = g(x) - h(x)$ and suppose that $\frac{d^n}{dx^n} f(x)$ exists in an interval containing the points $x_0, x_1, \cdots, x_n$. Then the error in interpolation may be estimated from the well-known relation

$$f(x_0) = \frac{f^{(n)}(\xi)}{n!} (x_0 - x_1)(x_0 - x_2) \cdots (x_0 - x_n), \tag{1}$$

where ξ is some point in the smallest interval containing $x_0, x_1, \cdots, x_n$.

In the most usual case, where $h(x)$ is a polynomial of degree less than n, we have $f^{(n)}(\xi) = g^{(n)}(\xi)$.

It is natural to consider the corresponding situation for functions of two independent real variables x and y. Let g and h be two functions such that $g(x, y) = h(x, y)$ for n points $x = x_i, y = y_i (i = 1, 2, \cdots, n)$. Setting $f(x, y) = g(x, y) - h(x, y)$ as before, we have $f(x_i, y_i) = 0$ for $i = 1, 2, \cdots, n$. Then if (x_0, y_0)

is a point at which g and h are defined, we may ask whether there is any formula corresponding to (1) from which the error $f(x_0, y_0)$ can be estimated.

Some restrictions must be placed upon the function f if any interesting results are to be obtained. Let us suppose that $f(x, y)$ can be expanded in a Taylor series about each of the points $(x_i, y_i)(i = 0, 1, \cdots, n)$ with a region of convergence sufficient to include all the points of the set. These conditions are more stringent ones than will be required for obtaining the later results; on the other hand, they would almost always be satisfied in any practical problem of interpolation, so it scarcely seems worthwhile to look for the weakest possible conditions at this point.

The first case of real interest is $n = 3$. It follows from the general statement of Taylor's theorem with the remainder that

$$
\begin{aligned}
0 = f(x_i, y_i) &= f(x_0, y_0) + (x_i - x_0) f_x(x_0, y_0) + (y_i - y_0)f_y(x_0, y_0) \\
&+ \tfrac{1}{2}[(x_i - x_0)^2 f_{xx}(\xi_i, \eta_i) + 2(x_i - x_0)(y_i - y_0)f_{xy}(\xi_i, \eta_i) \\
&+ (y_i - y_0)^2 f_{yy}(\xi_i, \eta_i)] \qquad (i = 1, 2, 3),
\end{aligned} \tag{2}
$$

where (ξ_i, η_i) is a point on the line segment joining (x_0, y_0) and (x_i, y_i) for $i = 1, 2, 3$.

The equation (2) may be regarded as a set of three linear equations in the two quantities $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$. The condition that these shall be consistent is

$$
\begin{vmatrix} f(x_0, y_0) + U_1 & x_1 - x_0 & y_1 - y_0 \\ f(x_0, y_0) + U_2 & x_2 - x_0 & y_2 - y_0 \\ f(x_0, y_0) + U_3 & x_3 - x_0 & y_3 - y_0 \end{vmatrix} = 0, \tag{3}
$$

where

$$
U_i = \tfrac{1}{2}[(x_i - x_0)^2 f_{xx}(\xi_i, \eta_i) + 2(x_i - x_0)(y_i - y_0)f_{xy}(\xi_i, \eta_i) + (y_i - y_0)^2 f_{yy}(\xi_i, \eta_i)] \qquad (i = 1, 2, 3).
$$

If the three points (x_i, y_i) $(i = 1, 2, 3)$ are not in a straight line, (3) can be written in the form

$$
f(x_0, y_0) = -\frac{\begin{vmatrix} U_1 & x_1 - x_0 & y_1 - y_0 \\ U_2 & x_2 - x_0 & y_2 - y_0 \\ U_3 & x_3 - x_0 & y_3 - y_0 \end{vmatrix}}{\begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}}. \tag{4}
$$

This expression is analogous to (1), though far less simple and elegant in form.

A similar treatment can evidently be used in all cases of the type $n = \dfrac{m(m + 1)}{2}$.

For example, for $n = 6$ the equation corresponding to (4) is

$$(5)\qquad f(x_0, y_0) = -\frac{\begin{vmatrix} V_1 & x_1 - x_0 & y_1 - y_0 & (x_1 - x_0)^2 & (x_1 - x_0)(y_1 - y_0) & (y_1 - y_0)^2 \\ V_2 & x_2 - x_0 & y_2 - y_0 & (x_2 - x_0)^2 & (x_2 - x_0)(y_2 - y_0) & (y_2 - y_0)^2 \\ V_3 & x_3 - x_0 & y_3 - y_0 & (x_3 - x_0)^2 & (x_3 - x_0)(y_3 - y_0) & (y_3 - y_0)^2 \\ V_4 & x_4 - x_0 & y_4 - y_0 & (x_4 - x_0)^2 & (x_4 - x_0)(y_4 - y_0) & (y_4 - y_0)^2 \\ V_5 & x_5 - y_0 & y_5 - y_0 & (x_5 - x_0)^2 & (x_5 - x_0)(y_5 - y_0) & (y_5 - y_0)^2 \\ V_6 & x_6 - x_0 & y_6 - y_0 & (x_6 - x_0)^2 & (x_6 - x_0)(y_6 - y_0) & (y_6 - y_0)^2 \end{vmatrix}}{\begin{vmatrix} 1 & x_1 & y_1^2 & x_1 & x_1y_1 & y_1^2 \\ 1 & x_2 & y_2^2 & x_2 & x_2y_2 & y_2^2 \\ 1 & x_3 & y_3^2 & x_3 & x_3y_3 & y_3^2 \\ 1 & x_4 & y_4^2 & x_4 & x_4y_4 & y_4^2 \\ 1 & x_5 & y_5^2 & x_5 & x_5y_5 & y_5^2 \\ 1 & x_6 & y_6^2 & x_0 & x_6y_6 & y_6^2 \end{vmatrix}}$$

where

$$V_i = \tfrac{1}{6}[(x_i - x_0)^3 f_{xxx}(\xi_i, \eta_i) + 3(x_i - x_0)^2(y_i - y_0)f_{xxy}(\xi_i, \eta_i) \\ +3(x_i - y_0)f_{xyy}(\xi_i, \eta_i) + (y_i - y_0)^3 f_{yyy}(\xi_i, \eta_i)] \quad (i = 1, 2, \cdots, 6).$$

(Equation (5) breaks down only if the six points $(x_1, y_1) \cdots (x_6, y_6)$ lie on a single conic.)

As an example of the general case we may consider $n = 4$. We write

$$f(x_i, y_i) = f(x_0, y_0) + (x_i - x_0)f_x(x_0, y_0) + (y_i - y_0)f_y(x_0, y_0) \\ +\tfrac{1}{2}[(x_i - x_0)^2 f_{xx}(\xi_i, \eta_i) + 2(x_i - x_0)(y_i - y_0)f_{xy}(\xi_i, \eta_i) \\ + (y_i - y_0)^2 f_{yy}(\xi_i, \eta_i)] \qquad (i = 1, 2, 3, 4).$$

Now,

$$f_{xx}(\xi_i, \eta_i) = f_{xx}(x_0, y_0) + (\xi_i - x_0)f_{xxx}(\xi_i', \eta_i') + (\eta_i - y_0)f_{xxy}(\xi_i', \eta_i'),$$

where (ξ_i', η_i') is a point on the line segment between (x_0, y_0) and (ξ_i, η_i). Proceeding as before yields

$$(6)\qquad f(x_0, y_0) = -\frac{\begin{vmatrix} W_1 & x_1 - x_0 & y_1 - y_0 & (x_1 - x_0)^2 \\ W_2 & x_2 - x_0 & y_2 - y_0 & (x_2 - x_0)^2 \\ W_3 & x_3 - x_0 & y_3 - y_0 & (x_3 - x_0)^2 \\ W_4 & x_4 - x_0 & y_4 - y_0 & (x_4 - x_0)^2 \end{vmatrix}}{\begin{vmatrix} 1 & x_1 & y_1 & x_1^2 \\ 1 & x_2 & y_2 & x_2^2 \\ 1 & x_3 & y_3 & x_3^2 \\ 1 & x_4 & y_4 & x_4^2 \end{vmatrix}}$$

with

$$W_i = \tfrac{1}{2}[(x_i - x_0)^2(\xi_i - x_0)f_{xxx}(\xi_i', \eta_i') + (x_i - x_0)^2(\eta_i - y_0)f_{xxy}(\xi_i', \eta_i')$$
$$+ 2(x_i - x_0)(y_i - y_0)f_{xy}(\xi_i, \eta_i) + (y_i - y_0)^2 f_{yy}(\xi_i, \eta_i)].$$

Corresponding formulas can be derived in this way for any value of n; in fact, several alternatives may be obtained in each case. In all cases the error $f(x_0, y_0)$ is given in terms of the derivatives of g alone if a polynomial of a certain type is used for the interpolating function. For equation (4), the suitable polynomial would be $h(x, y) = a + bx + cy$, for (5), $h(x, y) = a + bx + cy + dx^2 + exy + fy^2$; for (6), $h(x, y) = a + bx + cy + dx^2$. If the interpolating function $h(x, y)$ is not so chosen, the formulas remain valid, but derivatives of h will appear.

The same procedure is applicable to functions of any number of independent variables.

ON A LEMMA BY KOLMOGOROFF

By Kai-Lai Chung

Princeton University

The following lemma was proved by Kolmogoroff [1]:

If $e_1, e_2, \cdots, e_n$ are independent events and U an arbitrary event such that ($W(X)$ denoting the probability of X and $W_e(X)$ the conditional probability of X under the hypothesis of e)

$$W_{e_k}(U) \geqq u, \qquad W(e_1 + \cdots + e_n) \geqq u.$$

Then

$$W(U) \geqq \tfrac{1}{9}u^2.$$

This result seems of some interest in itself and may also have practical applications, for it is easily seen that [2] in general if $e_1, e_2, \cdots, e_n$ are arbitrary no information about $W_{e_1 + \cdots + e_n}(U)$ can be obtained from that about $W_{e_k}(U)$, $k = 1, \cdots, n$. From this point of view the constant 1/9 is interesting, though it is unimportant in Kolmogoroff's proof of the law of large numbers. Using his original method this constant can easily be improved to 1/8. However, the following method will give a better result. At the same time we shall put it into a more general form.

Let

$$W_{e_k}(U) \geqq \alpha, \qquad \sum_{k=1}^{n} W(e_k) \geqq \beta.$$

Then we have for $1 \leqq k \leqq n$,

(1) $$W(U) \geqq W(U(e_1 + \cdots e_k)) = W(Ue_1 + \cdots + Ue_k).$$

Now a simple case of certain inequalities due to Bonferroni and Frechet [3] states that for arbitray events $E_1, \cdots, E_k$ we have

(2) $$W(E_1 + \cdots + E_k) \geqq \sum_{i=1}^{k} W(E_i) - \sum_{1 \leq i < j \leq k} W(E_i E_j).$$

Applying this to (1), we obtain

$$W(U) \geqq \sum_{i=1}^{k} W(Ue_i) - \sum_{1 \leq i < j \leq k} W(Ue_i e_j)$$

$$\geqq \sum_{i=1}^{k} W(e_i) W_{e_i}(U) - \sum_{1 \leq i < j \leq k} W(e_i) W(e_j),$$

using the independence of $e_1, \cdots, e_k$. Hence

$$W(U) \geqq \alpha \sum_{i=1}^{k} W(e_i) - \frac{1}{2}\left(\sum_{i=1}^{k} W(e_i)\right)^2 + \frac{1}{2} \sum_{i=1}^{k} W^2(e_i).$$

By Cauchy's inequality,

$$\sum_{i=1}^{k} W^2(e_i) \geqq \frac{1}{k}\left(\sum_{i=1}^{k} W(e_i)\right)^2.$$

Writing $\sum_k = \sum_{i=1}^{k} W(e_i)$, we have

(3) $$W(U) \geqq \left[\alpha - \left(\frac{1}{2} - \frac{1}{2k}\right) \sum_k\right] \sum_k .$$

Now let $0 < \gamma < \gamma_0 \leq 1$ where γ and γ_0 are to be determined later. If there is an e_i, $1 \leq i \leq n$ such that $W(e_i) \geqq \gamma\beta$, then

(4) $$W(U) \geqq W(Ue_i) = W(e_i) W_{e_i}(U) \geqq \gamma\alpha\beta$$

If every $W(e_i) < \gamma\beta$, we determine $k(> 1)$ such that

$$\Sigma_{k-1} < \gamma_0\beta \leqq \Sigma_k ;$$

thus

$$\Sigma_k < \Sigma_{k-1} + \gamma\beta < (\gamma_0 + \gamma)\beta.$$

And (3) yields

(5) $$W(U) \geqq \left[\alpha - \frac{1}{2}\left(1 - \frac{1}{k}\right)(\gamma_0 + \gamma)\beta\right] \gamma_0\beta.$$

Now we choose γ so that the last terms in (4) and (5) be equal. This gives

$$\gamma = \frac{2\alpha - \left(1 - \frac{1}{k}\right)\gamma_0\beta}{2\alpha + \left(1 - \frac{1}{k}\right)\gamma_0\beta}\gamma_0.$$

To maximize γ, we put $\dfrac{d\gamma}{d\gamma_0} = 0$ and find

$$\gamma_0 = \frac{2(\sqrt{2} - 1)\alpha}{\beta}.$$

If $2(\sqrt{2} - 1)\alpha \leqq \beta$, this choice of γ_0 is admissible, and we obtain

$$\gamma = \frac{2 - \sqrt{2} + \frac{1}{k}(\sqrt{2} - 1)}{\sqrt{2} - \frac{1}{k}(\sqrt{2} - 1)} \frac{2(\sqrt{2} - 1)\alpha}{\beta}.$$

Thus we get (the first inequality being retained for small values of n)

(6) $$W(U) \geqq \frac{2 - \sqrt{2} + \frac{1}{n}(\sqrt{2} - 1)}{\sqrt{2} - \frac{1}{n}(\sqrt{2} - 1)} 2(\sqrt{2} - 1)\alpha^2$$
$$\geqq 2(\sqrt{2} - 1)^2\alpha^2 > \tfrac{34}{100}\alpha^2.$$

In case $2(\sqrt{2} - 1)\alpha > \beta$, we choose $\gamma_0 = 1$, and we obtain

$$\gamma = \frac{2\alpha - \left(1 - \frac{1}{k}\right)\beta}{2\alpha + \left(1 - \frac{1}{k}\right)\beta}.$$

Thus we get

$$W(U) \geqq \frac{2\alpha - \left(1 - \frac{1}{n}\right)\beta}{2\alpha + \left(1 - \frac{1}{n}\right)\beta}\alpha\beta$$
$$\geqq \frac{2\alpha - \beta}{2\alpha + \beta}\alpha\beta.$$

If we write $\beta = \eta\alpha$, we have

(7) $$W(U) \geqq \frac{2 - \eta}{2 + \eta}\eta\alpha^2.$$

We summarize (6) and (7) in the following table:

β/α	$\geqq 2(\sqrt{2} - 1)$	$= \eta < 2\sqrt{2} - 1$
$W(U)$	$\geqq 2(\sqrt{2} - 1)^2\alpha^2$	$\geqq \dfrac{2 - \eta}{2 + \eta}\,\eta\alpha^2$

Thus for Kolmogoroff's case ($\eta = 1$) we have $W(U) \geqq \frac{1}{3}\alpha^2$.

REFERENCES

[1] A. Kolmogoroff, "Bemerkungen zu meiner Arbeit 'Über die Summen zufälliger Grössen'," *Math. Annalen*, Vol. 102 (1929), pp. 434–488.

[2] K. L. Chung, "On mutually favorable events," *Annals of Math. Stat*, Vol. 13 (1942), pp. 338–349.

[3] M. Fréchet, *Les probabilitiés associeés à un système d'événements compatibles et dépendents*, Première partie, Hermann, Paris, 1939, p. 59.

APPROXIMATE WEIGHTS

By John W. Tukey

Princeton University

1. Summary. The greatest fractional increase in variance when a weighted mean is calculated with approximate weights is, quite closely, the square of the largest fractional error in an individual weight. The average increase will be about one-half this amount.

The use of weights accurate to two significant figures, or even to the nearest number of the form: 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 55, 60, 65, 70, 75, 80, 85, 90, or 95, that is to say, of the form $10(1)20(2)50(5)100 \times 10^r$ can thus reduce efficiency by at most $\frac{1}{4}$ percent, which is negligible in almost all applications.

2. Proof. Let the optimum weights be W_i, $i = 1, 2, \cdots, n$, with $W_i \geq 0$, where it is convenient to choose the normalization $\Sigma W_i = 1$. Let σ^2 be the variance of $\Sigma W_i x_i$, then the variance of each x_i must be σ^2/W_i, and since this is a weighted mean, the means of the x_i are the same.

Let the approximate weights be $W_i(1 + \lambda\theta_i)$, where $0 < \lambda < 1$ and $|\theta_i| \leq 1$, $i = 1, 2, \cdots, n$. Thus λ is the largest fractional error which may be made in the situation considered. We need the weak requirement $\lambda < 1!$ The approximately weighted mean is

$$\frac{\sum W_i(1 + \lambda\theta_i)x_i}{\sum W_i(1 + \lambda\theta_i)} = \sum W_i \frac{1 + \lambda\theta_i}{1 + \lambda\bar{\theta}},$$

where $\bar{\theta} = \Sigma W_i \theta_i$. Its variance is

$$\sum W_i^2 \left(\frac{1 + \lambda\theta_i}{1 - \lambda\bar{\theta}}\right)^2 \frac{\sigma^2}{W_i}$$

$$= \sigma^2 \left\{1 + \frac{\lambda}{1 + \lambda\bar{\theta}} \sum W_i(\theta_i - \bar{\theta}) + \frac{\lambda^2}{(1 + \lambda\bar{\theta})^2} \sum W_i(\theta_i - \bar{\theta})^2\right\}$$

$$= \sigma^2 \left\{1 + \lambda^2 \frac{(\sum W_i \theta_i^2) - \bar{\theta}^2}{(1 + \lambda\bar{\theta})^2}\right\},$$

and, since $\Sigma W_i \theta_i^2 \leq 1$, this is bounded by

$$\sigma^2 \left\{1 + \lambda^2 \frac{1 - \bar{\theta}^2}{(1 + \lambda\bar{\theta})^2}\right\}.$$

Now the only maximum of this expression for $|\bar{\theta}| \leq 1$ occurs when $\bar{\theta} = -\lambda$, and the bound becomes

$$\sigma^2 \left(1 + \frac{\lambda^2}{1 - \lambda^2}\right) = \frac{\sigma^2}{1 - \lambda^2}.$$

This proves the first statement in the summary.

The greatest fractional change which occurs when a number is approximated by one of the form $10(1)20(2)50(5)100 \times 10^r$ is 5/105, which occurs, for example, when 10.499999 $\cdots$, is replaced by 10. The same estimate applies to an approximation to two significant figures. The variance is thus multiplied by a factor bounded by

$$1 + \frac{5^2}{105^2 - 5^2} \leq 1.0023,$$

which proves the second statement.

The use of a weight of the simpler form 10, 15, 20, 30, 40, 50, 70, times a power of ten is seen in the same way to lead to an increase in variance and a decrease in efficiency of at most $4\frac{1}{4}$ percent.

3. Comment. It is interesting to compare the 90 possible values for 2 significant figures, the 35 possible values for the numbers proposed above, which might be called *two curtailed significant figures*, and the 24 possible values for logarithmic spacing at interval $(1.05)^2$, all of which extend over one power of ten with the same maximum fractional error in rounding. The use of the curtailed scheme for critical tables of weights and weighting coefficients would save more than 60 percent of the entries needed for two complete significant figures.

This device applies equally well to other numbers of significant figures.

ON THE USE OF THE NON-CENTRAL t-DISTRIBUTION FOR COMPARING PERCENTAGE POINTS OF NORMAL POPULATIONS

By John E. Walsh

Princeton University

1. Introduction. Consider two normal populations with the same variance and means μ and ν respectively. It is well known that confidence intervals and significance tests can be obtained for the difference $\mu - \nu$. Since μ is the 50% point of the first population and ν is the 50% point of the second population, this represents a particular solution of the general problem of obtaining confidence intervals and significance tests for the difference $\theta_\alpha - \varphi_\beta$, where θ_α is the α percent point of the first population and φ_β is the β percent point of the second population. The purpose of this note is to point out that the results of Johnson and Welch [1] for the non-central t-distribution can be used to furnish a solution of the general problem.

2. Analysis. Let A_γ be the γ percent point of the normal population with zero mean and unit variance (i.e. exactly γ% of the population has values less than A_γ). Then if σ is the common standard deviation,

$$\theta_\alpha = \mu + A_\alpha\sigma, \qquad \varphi_\beta = \nu + A_\beta\sigma$$

Thus

$$\theta_\alpha - \varphi_\beta = (\mu - \nu) + (A_\alpha - A_\beta)\sigma.$$

The non-central t-distribution investigated by Johnson and Welch in [1] is based on the quantity

$$t = (z + \delta)/\sqrt{\chi^2/f},$$

where z has a normal distribution with zero mean and unit variance, δ is a constant, and χ^2 has a χ^2-distribution with f degrees of freedom and is distributed independently of z. Methods and tables are given in [1] whereby a constant $t(f, \delta, \epsilon)$ can be computed having the property that

$$Pr[t > t(f, \delta, \epsilon)] = \epsilon.$$

These relations will be used to obtain confidence intervals for $\theta_\alpha - \varphi_\beta$. The resulting confidence intervals can be used to obtain significance tests for $\theta_\alpha - \varphi_\beta$.

Let $x_1, \cdots, x_n$ be a random sample of size n from the first population while $y_1, \cdots, y_m$ is a random sample of size m from the second population. Then consider

$$\frac{\bar{x} - \bar{y} - (\theta_\alpha - \varphi_\beta)}{\sqrt{\sum_1^n (x_i - \bar{x})^2 + \sum_1^m (y_i - \bar{y})^2}} \cdot \sqrt{\frac{m + n - 2}{\frac{1}{n} + \frac{1}{m}}}$$

$$= \frac{\left[\dfrac{\bar{x} - \bar{y} - (\mu - \nu)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}}\right] + \dfrac{(A_\beta - A_\alpha)}{\sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\dfrac{\sum_1^n (x_i - \bar{x})^2 + \sum_1^m (y_i - \bar{y})^2}{\sigma^2(m + n - 2)}}}.$$

This quantity has a non-central t-distribution with

$$\delta = (A_\beta - A_\alpha) \Big/ \sqrt{\frac{1}{n} + \frac{1}{m}}, \qquad f = m + \quad .$$

For notational simplicity let

$$t\left(m + n - 2, \frac{A_\beta - A_\alpha}{\sqrt{\frac{1}{n} + \frac{1}{m}}}, \epsilon\right) = t(\epsilon), \quad \sum_1^n (x_i - \bar{x})^2 = S_1^2, \quad \sum_1^m (y_i - \bar{y})^2 = S_2^2.$$

Then one-sided confidence intervals for $\theta_\alpha - \varphi_\beta$ with confidence coefficient ϵ are given by

$$\theta_\alpha - \varphi_\beta < \bar{x} - \bar{y} - \frac{t(\epsilon)\sqrt{S_1^2 + S_2^2}}{\sqrt{(m + n - 2) \Big/ \left(\frac{1}{n} + \frac{1}{m}\right)}},$$

$$\theta_\alpha - \varphi_\beta > \bar{x} - \bar{y} - \frac{t(1 - \epsilon)\sqrt{S_1^2 + S_2^2}}{\sqrt{(m + n - 2) \Big/ \left(\frac{1}{n} + \frac{1}{m}\right)}}.$$

Two-sided confidence intervals for $\theta_\alpha - \varphi_\beta$ with confidence coefficient

$$1 - (\epsilon_1 + \epsilon_2)$$

are given by

$$\bar{x} - \bar{y} - \frac{t(\epsilon_2)\sqrt{S_1^2 + S_2^2}}{\sqrt{(m + n - 2) \Big/ \left(\frac{1}{n} + \frac{1}{m}\right)}} < \theta_\alpha - \varphi_\beta < \bar{x} - \bar{y} - \frac{t(1 - \epsilon_1)\sqrt{S_1^2 + S_2^2}}{\sqrt{(m + n - 2) \Big/ \left(\frac{1}{n} + \frac{1}{m}\right)}},$$

where $\epsilon_1 + \epsilon_2 < 1$.

REFERENCE

[1] N. L. Johnson and B. L Welch, "Applications of the non-central t-distribution", *Biometrika*, Vol. 31 (1940), pp. 362-389.

THE TEACHING OF STATISTICS

A report of the Institute of Mathematical Statistics Committee on the Teaching of Statistics[1]

PREFATORY NOTE

This report on the teaching of statistics contains two parts. Part I is a summary of the conclusions reached by the committee concerning the appropriate content and organization of teaching in statistics. It is oriented towards the future, and is intended as a program for action. Part II, mainly the work of the chairman of the committee, is a more intensive discussion of the general problem. It surveys the present state of the teaching of statistics, probes some of the reasons for existing weaknesses in this teaching, and states more fully the basis for the conclusions summarized in Part I.

Additional material, with special reference to applied statistics, is contained in a report of The Committee on Applied Mathematical Statistics of the National Research Council, entitled *Personnel and Training Problems Created by the Recent Growth of Applied Statistics in the United States.*[2]

PART I

SUMMARY OF CONCLUSIONS

1. Who are the prospective students of statistics? A complete teaching program in statistics must be designed to meet the needs of four principal categories of students, listed here according to the amount of training in statistics that is needed to meet their requirements.

a. *All college students.* Statistical method is a vital branch of scientific method. It is widely used in most sciences, business, government, and ordinary life. Some understanding of the nature of inductive inference from quantitative data on the basis of the theory of probability as portrayed in statistical method is an indispensable part of a liberal education.

b. *Future consumers of statistics.* Some students will specialize in administration, business, or other subject-matter that will require them to understand the results of statistical analyses of special problems, although they themselves do not make these analyses. For example, business executives and government administrators must frequently base action on statistical studies. Research workers and teachers in many fields may not themselves use statistical methods, yet in order to keep abreast of their own or cognate fields they must read and understand studies using statistical methods.

c. *Future users of statistical methods.* A still smaller group of students of

[1] The Committee consists of Harold Hotelling, Chairman; Walter Bartky, W. Edwards Deming, Milton Friedman, and Paul Hoel.

[2] Copies may be obtained from the National Research Council, 2101 Constitution Ave., Washington 25.

statistics are training themselves for careers of specialization in economics, population, sociology, housing, business, business research, industrial design, industrial production, personnel, purchasing, public opinion, biology, agricultural science, metallurgy, physics, chemistry, psychology, or some other field that makes extensive use of statistics. Research in these fields often requires the use of advanced statistical techniques, and even the development of new statistical theory. Students planning to do such research need statistical theory and methods as a tool.

d. *Future producers and teachers of statistical methods.* The smallest, but in many respects most crucial group of students of statistics, are those who intend to specialize in statistical methods for the sake of statistical methodology. Many of these will become teachers or full-time research workers, though some will find posts in government and industry in high-grade statistical work, frequently requiring the development of new statistical theory and methods. These students will become tool-makers.

2. What should they be taught?

a. *All college students.*[3] The fundamental logic and philosophy of statistics can be taught at an early stage. It is perhaps an appropriate subject to include in the kind of survey courses of physical or social sciences that have become so common in recent years. Three or four weeks of lectures and discussions should suffice to acquaint the students with the broad principles of inductive inference. No mathematics need be included, although some elementary experiments may well be performed to instil the concepts of sampling variation, randomness, and statistical predictability. The student even at this stage can be made to recognize the fundamentally statistical character of most decisions, arising from the fact that they involve an element of uncertainty and a balancing of the importance of different types of errors. The student can be made to understand the fundamental difference between inductive and deductive statements, the nature of statistical estimation, and the nature of a statistical hypothesis. These concepts can be made concrete by illustrating them in terms of problems ranging from everyday questions such as whether to cross a street in the middle of the block on up to such vital problems as the construction of an appropriate social security plan, or the design of an efficient experiment for selecting the best variety of corn, or the selection of the best method of testing for the presence of a disease.

b. *Future consumers of statistics.* Future consumers of statistics need two kinds of training in statistics. First, they need some knowledge of the kind of statistical material available in their field of specialization; of the sources of such data; and of their limitations. To meet this need they require what may be called "descriptive statistics," which places special emphasis on their own field of specialization. A one-quarter or one-semester course in some department or division (e.g., in the social sciences, or biological sciences) should meet

[3] This recommendation is almost an exact parallel of one made by a committee on the teaching of statistics, appointed by the Royal Statistical Society and published by the Society in 1947 as a report to the Council, later published in the Journal of the Royal Statistical Society, vol. cx, Part I, 1947.

this need. In addition, they need a reasonably thorough understanding of what statistics can and cannot do, what the major statistical techniques are, and how to interpret the results obtained by the application of such techniques This need may be met for those students who have some mathematical background by all or part of the fundamental one-year course discussed in the next section. For students lacking this background, special courses along similar lines will be required.

c. *Future users of statistical methods.* It is essential for fruitful application that users of statistical methods should not mechanically apply procedures learned by rote or taken from a manual. Since few research problems fit perfectly into clearly defined patterns, nothing is so important to the successful collection and analysis of statistical data as adaptability and flexibility in using techniques. These require a thorough comprehension of the logical foundations of statistics, especially of the assumptions underlying its various technical devices, and sufficient knowledge of the derivations of these devices to be able to adapt them to the special circumstances that inevitably develop. To provide this background, a minimum of a full year fundamental course in statistical methods is essential, followed by courses of application. It is highly desirable that this fundamental course be based on calculus as a prerequisite, because without it a proper understanding of the development of statistical techniques cannot be attained. But this is probably impossible at present, in view of the unfortunately low level of mathematical training of most college students. As an expedient, and it is hoped a temporary expedient, it is recommended that the fundamental course be given in two sections, one requiring calculus, the other only a knowledge of first-year college algebra. A single course (or pair of courses, in line with the temporary expedient just mentioned) *should suffice for all departments*, because the core of statistical methods is common to all fields of study. Given in this way, the fundamental course can have the advantage of being taught by the most competent statisticians in the institution.

In addition to a thorough training in theory and methods, users of statistical methods need training in applications. This can be provided by courses in various applied fields. It is usually advisable that these courses be given in the department of application (agriculture, population, engineering, economics, psychology, etc.), and require the fundamental one-year course as a prerequisite.

d. *Future research workers and teachers of statistical method.* The future research workers and teachers of statistical method clearly require far more intensive training in theory than has so far been suggested. A fundamental prerequisite to such training is knowledge of some advanced mathematics. It is difficult to specify exactly what or how much mathematics is necessary, but something of the algebra of matrices and of the theory of functions are minimum necessities, and a good deal of additional knowledge of algebra, geometry, and analysis add richness and power to the work of the statistical theorist.

In addition to advanced mathematics and advanced work in statistical method, the future statistical theorist needs a good deal of work on applications, in the form either of experience or courses. He will be a tool-maker, and needs to

know by personal experience something of the problems of those who use his tools. One satisfactory arrangement is an internship in statistical research, as is currently provided by some institutions. By this arrangement, interns work under competent leadership in various government or private agencies that are engaged in large-scale statistical studies. The interns do research in theory, adapt the physical circumstances to theory and vice versa, and have actual practice in the design of experiments, the construction of questionnaires, writing of instructions, planning tabulations, analyzing the results, and examining sampling variances.

It is obvious that proper advanced courses in statistics will for many years be the province of a few institutions only, as there does not exist at present an adequate professional body to man more than a few.

3. Who should teach statistics? It is clear from the preceding section that two different kinds of courses are required to meet the needs of students of statistics: first, courses in statistical method and methodology; and second, courses in applications of statistical methods to particular fields.

The most important requirement for a successful university program in statistics is that courses in statistical method and methodology should be taught by a statistical theorist, a man who has had the training outlined in Art. 2d above, is specializing in statistics, is doing research in statistical method, and who has had some first-hand acquaintance with applications of statistical techniques. This is the only way such courses can be kept abreast of developments and sufficiently broad to meet the needs of all departments. This recommendation may seem to belabor the obvious, but a glance at the qualifications of most people currently teaching statistical methods will show why it is necessary.

Most courses in applications should be taught by people thoroughly conversant with the relevant subject-matter fields as well as statistical methodology. Some courses in applications may be taught by statistical theorists, particularly new applications or applications that are common to many fields.

4. How should the teaching of statistics be organized? The teaching program in statistics should be organized around a separate administrative unit, an Institute or Department of Statistics. This department should be primarily responsible for the teaching of courses in statistical methods: the fundamental course in statistical method described above, specialized methods for particular fields of application (e.g., factor analysis, time-series analysis), and advanced courses in statistical theory.

In addition, the department of statistics should offer its services as a consulting centre on problems in statistics arising in research in other departments of the institution, both as a service to these other departments and because research in statistical methods peculiarly requires stimulation from close communication with applications. The department of statistics might also provide laboratory facilities for itself and other departments,[4] and might undertake directly, or

[4] See the interesting suggestions on this point on p. 14 in *Personnel and Training Problems*, loc. cit.

through an associated research staff, special assignments involving the application of statistical methods to concrete problems.

Intermediate courses dealing primarily with applications ordinarily belong in other departments (agriculture, economics, demography, engineering, biology, etc.), although some may be given in the department of statistics. The exact location of courses in application will depend on the accident of the departmental affiliation of the persons competent to teach them. Coordination of the teaching program in statistics can be achieved by an interdepartmental committee. The department of statistics should not, however, consist of such a committee under a different name. It should be a thoroughly independent department, with all or most of its members entirely in the department.

The recommendation that the responsibility for teaching statistical methods be centered in a separate department is based on the belief that the teaching of statistical methods without theory can only be uninspiring and harmful; that a separate department of statistics offers the only arrangement that can assure statistical theory being taught by competent theorists, and the only satisfactory arrangement for ensuring the strong incentive for statistical research, with appropriate recognition and advancement, which is as necessary for the teaching of statistics as for the teaching of any other subject.

5. What should be done about adult education? The preceding recommendations are all directed toward the teaching of statistics to undergraduate and graduate students. There is an additional need that these do not meet, namely, the provision of training to mature research workers in various fields already established in their professions. This need arises in part from the inadequate teaching of statistics in the past, but even more from the extremely rapid advance in the theory and practice of statistics which have made it difficult for any but the specialist to keep abreast of developments. Some institutions are making efforts to meet this need by providing evening and late-afternoon classes for employed research workers. Such classes are feasible only in the larger centres of statistical activity. There is also the need of providing advanced research workers in particular fields with highly specialized guidance in selected topics. A department of statistics organized along the lines suggested above can contribute toward meeting this need by effective counseling of colleagues in other departments, and by organizing special seminars and lectures for them. The professional statistical associations are also contributing by arranging special expository programs.

PART II

THE PLACE OF STATISTICS IN THE UNIVERSITY[5]

Contents

A MINOR NUISANCES AND INEFFICIENCIES IN STATISTICAL TEACHING

6. Lack of coordination among departments. Lack of advanced courses and laboratory facilities. The teaching of statistics in American colleges and universities, which has for the most part been a development since the first world war and has now reached large proportions, presents a number of unsatisfactory features Courses in statistical methods are taught in various departments without coordination or inter-communication. These courses cover what is to a large extent the same material, but with many variations in the selection of subjects according to the ideas and abilities of individual instructors, and with

[5] An earlier version of this part, prepared entirely by the chairman, is being published by the University of California Press in a report of a symposium on probability and statistics. The Committee as a whole made and adopted the present condensation, with W. Edwards Deming and Milton Friedman contributing most of it. Publication of the Berkeley symposium, including the more detailed original, has been delayed, but it is expected to appear soon.

illustrative examples drawn in each case from material pertaining to the department in which the course is taught. Thus a student desiring to learn more about statistics than he can obtain in one department must, in taking courses in other departments, repeat a great deal of what he has previously covered.

There is a plethora of elementary courses and a dearth of advanced ones. Some departments have excellent statistical laboratories which they reserve for the use of their own students, each with an attendant to keep others away, while other departments have none. Some classes in elementary statistics are too large and some too small, with no one in a position to equalize the sections between different departments

7. Inefficient decentralization of libraries. The library situation is confused. Books on statistical methods are catalogued and shelved under Sociology, Economics, Business, Psychology, Zoology, Botany, Engineering, and Medicine. Books on probability are divided between Philosophy, Mathematics, Physics, and Chemistry. Books on the method of least squares are for the most part divided between Mathematics, Astronomy, and Civil Engineering, though some get into the Economics, Geology, and Physics reading-rooms. Works on the analysis of variance and design of experiments are likely to be concentrated under Agriculture, while methods of approximate evaluation of multiple integrals and similar purely mathematical subjects of use in statistics are, at least in one of our largest universities, to be found only in the library of Biology

B. THE MAJOR EVIL: FAILURE TO RECOGNIZE STATISTICAL METHOD AS A SCIENCE, REQUIRING SPECIALISTS TO TEACH IT

8. Too many teachers not specialists. The above nuisances are but minor. The major evil is that those attempting to teach statistical method are all too often not specialists in the subject. Their original selection was seldom on the basis of scholarship in this field; they are not encouraged to make advanced studies in it; and their environment is such as to draw their attention in every direction except to the central truths and problems of their science. Frequently they lack the knowledge of mathematics necessary to begin to read the more serious literature of the subject that they are teaching. Many have been utterly unable to keep up with the rapid progress which has been taking place in statistical methods and theory, *progress which affects even the most elementary things to be taught.*

9. Results: students ill equipped. There results a widespread teaching of wrong theories and inefficient methods. Students are sent to the government service and to industrial and commercial statistical positions equipped with the skill that results from careful drilling in methods that ought never to be used. Some of these same students are encouraged and assisted to become college and university teachers of statistics without ever making thorough-going studies of the fundamentals of the subject, or exhibiting any power of making original contributions to it, or studying any graduate mathematics. Through the method of selection of teachers in general use, and through textbooks written by individuals of this type, there is a perpetuation of obsolete ideas and unsound methods.

All this does not mean that any considerable number of people teaching statis-

tics are unworthy or objectionable members of the academic community. Many, indeed, are of superior intellect, upright character, personal charm, and undoubted teaching ability. Some are making creative contributions to other subjects. The only trouble is that they are teaching a subject in which they are not specialists, and which progresses so fast that only specialists can keep up with it.

10. Reasons why teachers of statistics are often not specialists. The chief reasons for the extensive teaching of statistical method by people who are not specialists in it appear to be the following:

a. *The rapid growth of the subject* and multiplication of its applications, creating a very large and very urgent demand for teaching it that could not be met immediately by the small existing number of scholars specializing in statistical method. This difficulty is aggravated by the paucity of university facilities for training advanced scholars in the field, so that even now the available number of such scholars cannot be expanded with sufficient rapidity to meet the current need. As specialists have not been available in anything like sufficient numbers, statistical method has inevitably been taught largely by non-specialists.

b. *Confusion between statistical method and applied statistics.* Statistical method is a coherent, unified science. "Applied statistics" may mean any of thousands of diverse things. Any particular study in applied statistics will ordinarily utilize some few of the results obtained by the science of statistical method, but will be largely concerned with matters peculiar to the particular application in view and others closely related to it. For example, studies of business cycles utilize statistical methods, good or bad, with a view to drawing inferences from existing data on prices, production, incomes, interest rates, bank reserves and the like. The main job of the applied statistician in this field is to study the sources and nature of the various series of observations, keeping in mind incidental events which may break the continuity of a series, and watching, with a background of economic theory and knowledge of the facts, for explanations He should also be well acquainted with statistical theory, since otherwise there is grave danger of wasting or misinterpreting the laboriously accumulated observations. Indeed, an organization studying business cycles, or solar cycles, or rat psychology or cancer or practically anything else, would almost certainly benefit from participation by a specialist in statistical method.

However, the chief attention in any such study will not be on statistical method but on features peculiar to its own scope. The specialist in statistical method will do well to participate occasionally in such a study, but if he does so too extensively the needs of the application will so engross his attention that he cannot keep up with the progress of statistical method itself.

The call of applications is enticing, and has led many young scholars to forsake the cultivation of statistical theory. The applications have benefited greatly by the process Moreover, problems brought back in this way from applications have provided valuable inspiration in developing theory. The mistake lies in supposing that participation in applied statistics is equivalent to specialization in statistical method and theory, and the consequent appointment to teach the latter of persons whose sole concern is with the former.

c. *Failure to recognize the need for continuing research* in the theory of statistics by those who teach it. There is an easy tendency to assume that all the requisite ideas and formulae can be found in some book, and that the duty of the teacher of statistics is simply to transfer this established book-knowledge to the minds of the students and impart to them skill in applying it. Similar attitudes applied to other subjects have in the past been a drag on progress, and have long been discarded in respectable universities. They still hang on, however, even in the best institutions with respect to statistics. The spectacular advances of the last three decades in statistics should make it clear to anyone who has followed them that statistical method is far from static, that the best techniques of present-day statistics may tomorrow be replaced by something better, and that unsolved problems regarding the theory and methods of statistics are sticking out in every direction. A vast amount of research, mostly of a highly mathematical character, is needed and is in prospect. Anyone who does not keep in active touch with this research will after a short time not be a suitable teacher of statistics. Unfortunately, too many people like to do their statistical work as they say their prayers—merely substitute in a formula found in a highly respected book written a long time ago.

d. *The system of making appointments to teach statistics within particular departments that are devoted primarily to other subjects.* In effect, the teacher of statistical method is too often selected by economists or sociologists or engineers or psychologists or medical men because he is to teach in one of these departments. Thus the task of selection devolves upon people unacquainted with the subject, though realizing the need for it in connection with a very specific application. Under such conditions there is an inevitable tendency to emphasize the immediately practical and specific at the expense of the fundamental work of wider applicability and greater long-run importance. Confusion between a science and its applications is most pronounced with those who know little about it, and the distinction between statistical method and applied statistics is likely to be completely lost when a sociologist or an engineer is confronted with the problem of finding someone to teach statistics. If he does make the distinction at all he is likely to choose in favor of applied statistics.

Strangely, the actual teaching that ensues is bound to consist largely of statistical theory, because the students will ordinarily not have had statistical theory elsewhere, and they must have some in order to apply it. What often happens is that a sociologist or an engineer who has made some study of statistics embarks on what he thinks will be a career of teaching the application of statistical method to sociological or engineering problems, only to discover that because of the ignorance of the students he is compelled to teach the fundamentals of statistics, an entirely different subject for which he lacks preparation, talent, and interest.

An incident of this sort has been cited previously.[6] A prominent economist was asked to teach a course entitled "Price forecasting" in a leading university, and accepted. He found, however, that his lectures on this subject were over

[6] Harold Hotelling, "The teaching of statistics" *Annals of Math Stat.*, vol. xi, 1940, pp 457–470.

the heads of the students because he was using statistical concepts unfamiliar to them. He therefore went back over the ground covered so as to explain these particular statistical concepts along with their application. But in explaining them he found himself using other statistical concepts, which in turn called for explanation. At the end of the semester he found that he had not given the course in price forecasting which he had planned, and for which the large class had enrolled, but instead had taught a somewhat disordered course in elementary statistics, a subject in which he did not feel particularly competent, and for which the students had not come. When he was asked to teach price forecasting a year later he proposed that a prerequisite of a course in statistics be imposed, but this proposal was rejected by the chairman of the department, and the course was not repeated

11. Appointments under the existing system are not all bad. More by accident than by design in the existing system, not all statistical appointments by departments of application are bad. Some professors in these departments make conscientious excursions into statistical theory, are well advised by competent specialists in statistics, and bring about the appointment of men of high quality well acquainted with statistical method and theory of the currently best sort. This may work out well if the man so appointed is an able and energetic scholar deeply devoted to his subject, if he is placed immediately in the highest professorial rank, and if he does not feel under obligation to devote himself too exclusively to the special interests of the department of which he finds himself a member. He is then free to pursue his specialty, to keep informed on the latest developments in statistical method and himself to add to the subject, while at the same time transmitting to students a well rounded and up-to-date selection of knowledge. It is in this way that some of the present leaders in statistics have developed It is a wrong procedure, however, to depend on accidents of this kind.

The system of departmental organization and of making appointments and recognizing proficiency in the teaching of statistics needs to be altered. The usual story is typified by the appointment of a promising young scholar in statistical method to a junior position in some department of application where he is expected to work on problems and to teach statistical methods with a sole eye to the work of the specific department. He is then under pressure to concentrate on a particular kind of applied statistics, for his advancement will depend, not on his statistical attainments at all, but on his study of the literature, terminology, techniques and theories of the application. His usual associates will be in the department in which he is teaching rather than others teaching statistics. The loss, although not total, is great, because the opportunity to make the most of the man's statistical ability is lost, and his ability as an economist, agricultural scientist, engineer, or something else that he is not particularly fitted for, is substituted

A still less favorable circumstance, and unfortunately more common, is that in which the teacher of statistics is not even selected for scholarship in the theory of statistics. Studies in some other field, with some slight dabbling in the appli-

cation of statistical methods to it, plus a pleasing personality, have all too frequently been thought to comprise sufficient qualifications for teaching statistical methods and theory.

12. Unsatisfactory texts. The uncritical character of the teaching is reflected in the long line of textbooks written by teachers who have not made any genuinely fundamental study of statistics, but pass on to students in a magisterial fashion what was passed on to them. Authority takes the place of derivations and ultimate sources It is no wonder that these textbooks, copied from each other, contain increasing accumulations of errors, or that long delays have intervened between the introduction of important new statistical methods and theories in the periodical literature and their appearance in the textbooks and courses put before students.

The latest discoveries in the theory of statistics affect what should be taught in elementary courses, and no syllabus can be expected to survive more than a few years of research. The development of new statistical methods and ideas of overwhelming importance must be allowed to compete with material already well established as true and useful. The new material is equally true and in some cases even more useful than matter usually incorporated in the best of current courses and textbooks.

13. Omission of probability theory from texts and teaching. One of the important weaknesses in much of the current teaching of statistics is a failure to make proper use of the theory of probability. Without probability theory, statistical methods are of only minor value, for although they may put data into forms from which intuitive inferences are easy, such inferences are very likely to be incorrect. The objective weighing of the degree of confidence to be placed in inductive conclusions is necessary to avoid fallacies. Indeed, the whole foundation of descriptive statistical methods, of inductive inference, and of the design of experiments, rests upon probability theory.

The relevance of probability to much statistical work was indeed questioned a quarter-century ago by a group of economists impressed by the lack of independence between consecutive observations, and this attitude, in conjunction with an exaggerated and belated remnant of nineteenth-century empiricism, has had a certain influence, particularly on the statistical methods in use by economists. This view is now rapidly giving way to a tendency to use the powerful new statistical methods discovered in the meantime. It is now perceived that efficient objective methods can be used over a much wider range of cases than was formerly supposed, because the independence assumed in their derivations refers not to observations but to residuals from the theoretical model used. Furthermore, research is under way, and has already achieved promising results, on the extension of accurate methods to still more extensive classes of problems.

C. PROPER QUALIFICATIONS OF TEACHERS OF STATISTICS

14. Statistics compared with other subjects. The qualifications appropriate for teachers of statistical method and theory are not essentially different in degree from those for teachers of other subjects in the same institutions; proficiency in

statistical method and theory is merely to be substituted for it in other subjects. This substitution is, however, vital. It must not be imagined that proficiency in some other subject in which statistical methods are used incidentally is equivalent to proficiency in statistical method itself. The error of such a supposition, if carried over into another field, might lead to the appointment of a man as professor of chemistry on the ground that he could cook.

The first requisite of the college or university professor of any subject is a profound and thorough knowledge of that subject. It is customary in the better institutions at least to restrict appointments to the rank of assistant professor to persons who have demonstrated scholarly qualifications by work equivalent to that leading to a Ph.D. degree, including an original contribution to the subject that the individual is to teach. Promotion to the higher ranks is conditioned upon a number of criteria, among which published research is by far the most important in those institutions.

15. Current research in statistical method is essential for teachers. Research is even more essential in the teacher of statistics than in teachers of most other subjects, because so much remains to be worked out that is of immediate importance. Some college teachers do no research. This is usually regarded as deplorable. The evil is, however, of quite different magnitude according to the nature of what is taught by such teachers. In a new subject in which sharp differences of opinion exist or have recently existed on fundamental questions, in which current discoveries have an important bearing, and in which there have not yet been the time and consensus necessary for the preparation of an adequate and virtually error-free textbook, teaching without research may have calamitous effects. The effective teacher must, of course, have teaching ability, but no skill in pedagogy, no lustre of personality, can atone for teaching errors instead of truth. Errors are very likely to be taught by those who do no research, and then the more skillful the pedagogic indoctrination, the greater the harm. Sound educational policy calls for devotion to research of a large fraction of the time and energies of the teaching staff in a subject like statistical theory. Students also are in particular need of encouragement to do original and critical work in relatively new areas of this kind. They must be taught to shun the use of formulae and methods given merely on authority without full and convincing reasons, and to insist on looking closely and critically at assertions.

Even in the teaching of elementary statistical methods for direct practical use by specific occupational groups, where it might be thought that the teaching would most predominate over the research element, the teacher must face difficult questions whose answers call for research in statistical theory. Let us illustrate this by one example out of the many possible. In teaching the analysis of variance for use in agricultural experimentation, questions arising out of the possible non-normality of the underlying distributions must be dealt with in some way The formulae, even those in the best textbooks, are accurate only if the distribution is normal, and neither this fact nor the non-normality of many distributions should be concealed from the students. Obviously something more

needs to be said on the subject at this point. What the teacher can say depends on how deep he has gone into a whole series of perplexing questions, on some of which the views of scholars are not yet stabilized, and on which a tremendous amount of research is needed before the maximum practical value can be attained for a technique whose usefulness is already amazing.

16. Minimum requirements in mathematics for the training of teachers and research men in statistical theory. Because research in the theory of statistics requires advanced mathematics, and is indeed largely mathematical in character, a mastery of a substantial amount of higher mathematics must be an essential part of the training of prospective teachers of statistics. To specify exactly what or how much mathematics is necessary would be a difficult task. Something of the algebra of matrices and of the theory of functions are minimum necessities, and a good deal of additional knowledge of algebra, geometry, and analysis add richness and power to the work of the statistical theorist, the inventor of new statistical methods On the other hand, the time of the graduate student in statistics is much occupied with the theory of statistics itself; and some of his time should also go into the study of applied statistics. If the students entering a graduate school for advanced work in statistics went there equipped with a knowledge of matrix algebra and theory of functions and some additional higher mathematics, as is obtainable by undergraduates at some institutions, they would have time for applied statistics and could do some real work on applications.

There is a cruel dilemma here, resulting from the delay in learning mathematics imposed by the elementary curricula which have become customary in this country. The weakness of the mathematical element in the prevailing curricula affects both teachers and students of statistics to an extent justifying some attention from those interested in the improvement of statistics. In American universities elementary calculus is not often taught before the sophomore year, and the more advanced parts of algebra come still later, if at all.

If calculus could be pushed down into the high schools and assumed as a prerequisite for college courses in mathematics, statistics, economics, physics and several other subjects, the efficiency of instruction in all these departments could be increased. For example the difficulties experienced by students of economics with ideas of marginal cost, marginal revenue and the like correspond closely with the difficulties experienced by mathematicians for centuries in trying to define infinitesimals and derivatives, but now successfully overcome. The student who really knows differential calculus need not have the slightest difficulty with the marginal ideas of economics. Similarly in physics, the fundamental concepts of speed, acceleration, potential theory, conductivity, thermal capacity and radiation, are all mathematical and easier to grasp once and for all as such than to be learned afresh with each new application from textbooks in physics sometimes not clearly written and taught by teachers who must for one reason or another avoid a mathematical approach.

The possibilities of teaching quite advanced mathematics to young children

have scarcely begun to be explored. Children of kindergarten age are fascinated and thrilled by the wonders of topology. Groups and number theory can be tremendous sensations in the fifth grade, though all these subjects are ordinarily reserved for graduate students specializing in mathematics. What is lacking is teachers who know mathematics and its applications and who possess enough freedom to teach what they know instead of the long, dull and relatively useless drill on problems of wallpaper hanging and the like, problems turning on mere conventions which are quickly forgotten, painful repetitious work which makes children resolve to quit mathematics as soon as possible.

D. NEED FOR RELATING THEORY WITH APPLIED STATISTICS

17. An example of the interaction between theory and practice. A professor of psychology working with mental tests might enlist the assistance of a young statistical theorist with mutual benefit. The young man might for a short time do some of the drudgery of scoring tests and computing, passing on soon to the problems of test construction and the distribution of various functions of correlation coefficients. This last is on a new and exciting frontier of statistical theory The advancement of this frontier, which is really the main business of the young man in his capacity as prospective statistical theorist, would in this way come to him naturally as a problem or series of problems having a tangible meaning additional to its mathematical content. The empirical context is in such cases often of great value in suggesting suitable approaches, for example, suitable approximations in the study of functions not susceptible to simple mathematical representation in terms of elementary functions.

If the young theorist succeeds in extending the boundaries of multivariate statistical analysis by discovering the distribution of some new function of correlation coefficients, the chances are that this discovery will also have applications in anthropology, medicine, banking, and other pursuits which in the aggregate will greatly outweigh the application originally in view.

The discovery should be regarded primarily as a contribution to the general theory of statistics, and published in a journal devoted to mathematical statistics. It will then become available to a wide circle of teachers of statistics, who may incorporate it into their courses, and its methods and results will be studied by other investigators from the standpoint of possible generalizations and analogs. The importance of the discovery would be much more limited if it were thought of as a development in psychology and published only in a psychological journal. Perhaps dual or multiple publication ought to be permitted in such cases, but the first publication should be in a journal of mathematical statistics. Far too many good statistical ideas have been buried in connexion with obscure special applications.

18. Supplying opportunities for applications in graduate studies of statistics. The statistician who does any work in applications must know statistics as an art as well as a science The theoretical statistician, if he wishes to be of the utmost use to his colleagues in other disciplines, needs to know by personal

experience something of their lives and collateral problems. Indeed, experience with applications, and the challenge of problems arising out of applications, have played a most important part in the development of statistical theory. It follows that the graduate student in statistics needs contact with applied statistics which the institution should undertake to provide, or at least facilitate This need is next in importance after the needs for theoretical statistics and for pure mathematics. The distribution of time among the three—theoretical statistics, mathematics, and applied statistics—is hard to specify exactly, and must in any case depend on the nature of the student's previous work. If his mathematical preparation has been full and rich, more time should be spent on applied statistics in his graduate years than if he has already had substantial contact with applied statistics in some other way but is deficient in higher mathematics.

Applied statistics entails a somewhat detailed acquaintance with the field of application. Such a field might be life insurance, or mental testing, or industrial quality control, or sampling in the work of the Bureau of the Census or some other government agency; it might be agricultural economics, or business cycles. Proficiency in any such field calls for rather prolonged study, and it would be too much to expect the embryo statistical theorist to reach this stage of advancement in all subjects. He should, however, make more than a superficial study of some chosen field of application. This study might or might not be at the university. The requisite familiarity with applied statistics might in some cases be acquired by work in a government bureau, or in a research organization studying business cycles or something else involving applied statistics. What is most desirable is that the work should have brought the student to the point both of applying statistical methods in a reasonably effective way, and of perceiving the limitations of existing statistical methods. Perception of existing limitations has frequently been the germ of progress in the subject.

One satisfactory arrangement is an internship in statistical research, as is currently provided by some institutions. By this arrangement, interns work under competent leadership in various government or private agencies that are engaged in large-scale statistical studies. The interns do research in theory, adapt the physical circumstances to theory and vice versa, and have actual practice in the design of experiments, construction of questionnaires, writing of instructions and tabulation plans, analysis of the results and appraisal of sampling variances.

E. RECOMMENDATIONS ON THE ORGANIZATION OF STATISTICAL TEACHING AND RESEARCH IN INSTITUTIONS OF HIGHER LEARNING

19. Research should be encouraged; teaching schedules should not be overloaded. Colleges and universities usually expect the members of their faculties to engage in research as well as in teaching, the relative emphasis on these two functions varying greatly from institution to institution and to a lesser extent among departments within the same institution. Reasons why teachers of

statistics must do current research in order to teach the subject have already been given in Art. 15. In the organization of statistical teaching it is thus of extraordinary importance that colleges and universities emphasize research in the theory of statistics as a leading part of the work of the teaching staff in this field. Hours of teaching and other duties must be kept within such bounds as to make research possible, the initial selection of teachers must be of persons capable of research in statistics, and there must be provision of needed secretarial, computational and other assistance. The library must be adequate, not only in publications containing statistical theory, but in the larger field of pure mathematics as well.

20. Organizing statistical service in the university. In addition to the customary duties of teaching and research, faculty members expert in statistical methods find that they cannot escape a third, viz., advice to their colleagues and others regarding the statistical aspects of their problems. This often takes a good deal of time. Clearly it is in the interest of the academic enterprise that such services be provided. Scholars in many departments are finding that their work is greatly improved by competent statistical advice not only in the interpretation of their data but also in the design of their experiments and other investigations. The provision of competent advice frequently requires extended consideration of the general content of the problem as well as special analysis of its statistical features. And initial advice often needs to be supplemented by further service. The statistician, like the physician, often finds that one interview at which a prescription is dispensed does not end the matter satisfactorily.

Teaching hours must be distinctly limited if statisticians are to be able to render this service to the rest of the institution as well as maintain a high level of research in their own field.

One way to handle the problem of statistical service, especially in a large institution, is through a special organization devoted to this purpose. Such an organization, whether called a Statistical Institute, a Department of Applied Statistics, Statistical Laboratory, or something else, might supply not only advice but a more active kind of assistance, including computational and chart-drawing services.

A statistical service organization should be removed from the teaching of statistics only to the extent necessary to gain the advantages of some degree of specialization and to prevent undue interruption of the teacher's other work of teaching and of research in theory. There are distinct advantages for all parties in a fairly close connexion between practical statistical work, research in statistical theory, and statistical teaching. Each of these activities benefits the others, provided only that it does not take away from it too much time. Research in statistical theory, like medical research, needs frequent revitalizing injections of specific practical problems. It also needs the stimulus of contact with students. The teaching of statistical method is made more vigorous both by research in the subject and by the presence of applications with which students can be confronted. And the needs of applications are better met if through an organiza-

tion such as is here envisaged they can be brought to the attention of appropriate specialists, and if also students can be enlisted when needed for their treatment.

A university organization dealing with statistics may properly comprise two parts with overlapping personnel, one devoted chiefly to applied statistics, the other to theoretical statistics. The teaching might be done by both, but at least at the more advanced levels would be primarily the concern of the theoretical part. Migration between the two ought to be easy and frequent, though some individuals are so definitely adapted to one kind of work or the other as to make it undesirable to have fixed rules calling for periodic transfers.

In smaller institutions it may not be practicable to have statistical organizations sufficiently well staffed to provide adequate consulting service. To meet the needs in some of these cases regional centres for advice and service in applied statistics might be established at large universities throughout the country, with access made readily available for sister institutions. These centres might also carry on work in applied statistics in behalf of government agencies and other organizations, much as various agricultural colleges have for years been carrying on cooperative work with the federal Department of Agriculture.

The question how far, if at all, such a university centre of applied statistics should go into the market place and engage commercially in service to business concerns is a debatable one. While there may be favorable reactions upon scientific work, there are grave dangers to the intellectual integrity of the institution which need serious consideration.

21. Organization for teaching. Passing from questions of personnel and the research and service functions of academic statisticians to teaching itself, we have to consider problems of departmental organization, of course contents, of systems of prerequisites, and of methods of teaching. All these we consider secondary problems, not in the sense of being unimportant, but because we believe that proper solutions of them will be reached with reasonable promptness when personnel of the kind described in Sec. C of this report are at work in some such general setting as has just been described. The ideas recorded below are general in character and are to be regarded as a starting-point for developing a program in a particular institution, once suitable faculty members have been obtained.

The teaching of statistics may be organized in any of the following ways:

a. In a department of theory and a department of applied statistics, both forming an Institute of Statistics.
b. In a single Department of Statistics.
c. Under an inter-departmental committee.
d. Under the exclusive jurisdiction of the Department of Mathematics.
e. It may be scattered among heterogeneous departments of application, without formal coordination.

Only a few large institutions will be in position to adopt the first plan. It is likely that the second will be most suitable for the majority. The third should probably be regarded as a makeshift for the transitional period until a proper department of statistics can be organized, a step that will not at the moment be

reasonably possible for most institutions because the right kind of scholarly personnel does not exist in adequate numbers. It is of course possible that some vestige of an inter departmental committee, perhaps in the form of an Advisory Board, might be a useful adjunct of a department of statistics in order to keep it informed of the needs of applications. It is also possible that something of the sort might function with respect to a department of mathematics, or any other department. On the other hand, the desired consultations and adjustments might be accomplished in less formal ways.

To make statistics a subdivision of a mathematics department is a solution that will appeal to administrators desirous of keeping down the number of departments. The subject-matter of statistics is to a sufficient extent mathematical to give some apparent weight to this plan, and some mathematicians have the unsound idea that any mathematician can teach statistics without specialized study or experience in application. On the other hand, statistics has some features uncongenial to traditional mathematics, arising partly from the urgency of practical needs which go beyond what can immediately be provided by rigorous mathematical theory. Again we may cite the problem in the teaching of the analysis of variance of what to do about possible non-normality of the underlying distribution (Art. 15). The user of this technique has the responsibility of verifying that the situation conforms to the assumptions, including that of normality, underlying the tabulated probability criteria. But he is in a very poor position to do this in a large proportion of the applications actually made of the analysis of variance. Yet the analysis of variance in some form—possibly through the use of rank-order numbers or through a transformation or some other auxiliary device—remains the one powerful means of attacking a very large and important class of practical situations. The practicing statistician needs to do some highly educated guessing on such matters—guessing that will be assisted but not made determinate by knowledge of a considerable range of mathematical truths regarding approaches to the normal distribution, moments of the variance-ratio in samples from non-normal populations, asymptotic large-sample theory, and other such topics. His mathematical insight needs to be supplemented by consideration of the particular subject-matter of application. Moreover, it is desirable that students of statistics have some practice with actual empirical data designed to develop the art of guessing in such ways.

Another example of non-rigorous mathematics used extensively in statistics is the whole business of asymptotic standard errors found by the differential method. It is desirable that good mathematics replace bad in such connexions, but something is to be said for the position into which so many practical statisticians have been driven, that even bad mathematics may be better than none at all. The requisite good mathematics along these lines can come only through those who have made really serious studies of statistics, though a sufficiently interested pure mathematician might eventually be led by such a student of statistics to undertake and complete the necessary research. Practical needs make approximations necessary; the goodness of a particular approximation can

often be judged adequately by a statistician familiar with the particular application long before the heavy artillery of advanced mathematical analysis can be brought up.

The teacher of statistics must have a genuine sympathy and understanding for applications, and these are not possessed by many pure mathematicians, at least in the opinion of some of those concerned with the applications; and it is this opinion rather than the possible fact that is of interest at the moment. For so long as such an opinion is maintained, for example by psychologists and economists, these specialists will be suspicious that courses in statistics given by a department consisting largely of pure mathematicians is unsuitable for their purposes. The result is likely to be a sabotaging of attempts at centralization, the different departments reverting to the old and ultimately objectionable system of teaching their own separate courses in statistical methods."

These difficulties are not necessarily insuperable, and it is to be expected that many medium-sized and small institutions will make their mathematical departments responsible for statistical teaching. But this ought not be be done without a consideration of the possible dangers.

22. The statistical curriculum. We next consider curricular problems These may be divided into those of the graduate school and those of the undergraduate college. Those of the graduate school may in turn be divided into those of specialization in statistics and of auxiliary teaching of statistics to students in other departments, such as sociology, who need to use statistical methods, have not studied them sufficiently as undergraduates, and cannot afford to put much time on them. Of these two subdivisions the number of students at present is greater in the second and the ultimate importance is greater in the first, because the whole future of statistics depends on improvement and enlargement of this graduate teaching.

The incidental teaching of elementary statistical methods to graduate students in other subjects, without any prerequisite in mathematics or statistics, cannot equip these students with a command of the subject at all comparable to that which could be obtained by a better integration of undergraduate with graduate work. A prospective sociologist, economist, psychologist, or physicist ought to study elementary statistical methods and concepts while still an undergraduate, and without special reference to his ultimate field of specialization.

The features of statistical methods peculiar in their applications, beyond what is taught through illustrations and exercises in an elementary course, may be fit material for a course, graduate or undergraduate, in a department of the application. Such a course should require as a prerequisite an elementary course in a department of statistics, or at least one taught by specialists in statistical method and theory.

For the undergraduate college, in place of the sporadic offerings now current in different departments, we recommend a combination of two general fundamental courses with a number of advanced courses. Of the latter some will be specialized to the work of particular departments or groups of departments.

Of the two fundamental courses one will require calculus as a prerequisite, the other only a knowledge of first-year algebra. It is to be hoped that the less mathematical of these two general statistical courses, instead of being elected by a majority of students, will gradually approach extinction, while the course based on calculus will become the vital point of contact of the student body with the concepts of statistics. The chief reason for insisting upon the importance of calculus as a prerequisite is simply the possibility of covering important statistical theory that is inaccessible to those who do not have it.

Modern statistical methods are based on the theory of probability. The general courses in statistics may therefore well begin with elementary probability. The duality between probability and statistical concepts,[7] for example between probability and relative frequency, between mathematical expectation and a sample mean, between parameter and statistic, should be explained. Derivations and the place of the normal distribution should be sketched, and the Student distribution should be derived and applied to a variety of problems in the first course based on calculus. Later courses given by the department of statistics, or whoever specializes in statistical theory, will naturally cover other statistical methods and theories. At the same time useful courses can be offered in economic statistics, mental testing, and other fields using statistical methods by specialists, regardless of departmental affiliation. There might be departmental cooperation; for example, the department of statistics might offer elementary and advanced courses in correlation and multivariate analysis, and the department of psychology might require these as prerequisites for some of its work in mental testing.

The teaching of statistics should be accompanied by considerable work in applied statistical problems, as well as exercises in mathematical theory, on the part of the students. A large part of this work in applied statistics is best conducted in a laboratory equipped with calculating machines, mathematical tables, drafting instruments, and other appurtenances.

Statistical laboratories require supervision, administration and maintenance. They are needed not only for the purpose of teaching statistics, pure and applied, at all levels, but also by research workers in many fields. There are possible gains of efficiency and economy in a centralized administration of them. One suggestion is that they be under the supervision of the university library. Another is that responsibility for them be lodged in a central department of statistics, or in a two-department statistical institute. Centralization can be carried too far, and it is likely that some units in a large organization will find it advantageous to have machines which are exclusively their own. The conflicting claims regarding machines and laboratories will require careful weighing.

23. Statistical method as a part of a liberal education. A question may also be raised as to whether some work in the statistical method should not be required of all college students as a part of a liberal education. This would be

[7] Cf. the article "Frequency distribution," *Encycl. of the Social Sciences* (1931).

a novel step, but has much to be said for it in view of the widespread use of statistics and growing interest in statistics.[5] Another point is that the student who can't make up his mind as to his ultimate field of specialization or vocation will do well to study those things that can be used in many fields. Of such things, mathematics and statistics are leading examples. There are more or less sound objections to systems of required studies, but if we are to have them, the claim of statistics should not be rejected merely on grounds of novelty.

ABSTRACTS OF PAPERS

Presented December 22, 1947 at the Berkeley Meeting of the Institute

1. The Performance Characteristic of Certain Methods for Obtaining Confidence Intervals. B. M BENNETT and J. NEYMAN, University of California, Berkeley.

Certain methods for obtaining confidence limits have been introduced by Bliss, R. A. Fisher and Paulson. Thus, e.g., let x_i, y_i $(i = 1, \cdots, n)$ represent a sample from a bivariate normal population with means $E(x_i) = \xi$, $E(y_i) = \alpha\xi$ and variances and covariance σ_x^2, σ_y^2, σ_{xy}. If $\bar{x}$, $\bar{y}$, S_x^2, S_y^2, S_{xy} are the sample means, variances and covariance respectively, then in order to determine confidence limits for α, the ratio:

$$u = \frac{\sqrt{n}(\bar{y} - \alpha\bar{x})}{\sqrt{S_y^2 - 2\alpha S_{xy} + \alpha^2 S_x^2}}$$

may be referred to the appropriate value t_ϵ of the Student-t distribution. The inequality: $| u | < t_\epsilon$ may, in general, be solved as a quadratic equation in α to yield two values $\underline{\alpha}$, $\bar{\alpha}$ which are presumed to be confidence limits for α. In this paper the probability π of being correct in using such a procedure, i.e., the performance or operating characteristic, is computed in the limiting case when σ_x^2, σ_y^2, $\sigma_{xy} = \rho\sigma_x\sigma_y$ are assumed to be known. It is shown that π is a function $\pi(\alpha, \xi, \sigma_x, \sigma_y, \rho)$ of all the parameters, and in particular of α itself, the quantity for which confidence limits are supposed to be provided. Similar "quadratic" methods are also used in certain regression problems, e.g., in determining confidence limits for a value of x corresponding to an additional value of y when a previous sample regression of y on x is available; or in determining confidence limits for the intersection point of two population regression lines. The performance characteristic of each of these methods is shown to be a function of the quantity for which the method gives confidence limits.

2. Some Further Results on the Bernoulli Process. T. E. HARRIS, Douglas Aircraft Co.

Let $z_1, z_2, z_3, \cdots$, be a sequence of random variables defined as follows: $P(z_1 = r) = p_r$, $r = 0, 1, 2, \cdots, k$. If $z_n = 0$, $z_{n+1} = 0$. If $z_n = r$, $r \neq 0$, then z_{n+1} is distributed as the sum of r independent random variables, each having the same distribution as z_1. It is assumed that $x < 1$, where $x = E(z_1)$. Let N be the smallest value of n such that $z_{n+1} = 0$. A method is given for obtaining an expansion of the moment-generating function of N. In the case where $p_r = 0$ for $r \geq 3$, this expansion takes the form $1 + (1 - e^{-s})(1 - p_0) F(s)$, where $F(s) = f_1(s) - p_2(1 - p_0)f_2(s) = 2xp_2^2(1 - p_0)^2 f_3(s) - \cdots$, where $f_1(s) = (e^{-s} - x)^{-1}$, and $f_n(s) = f_{n-1}(s)(e^{-s} - x^n)^{-1}$. Certain restrictions on the constants p_r insure that this expansion converges for a complex neighborhood of $s = 0$.

3. Most Powerful Tests of Composite Hypotheses I. Normal Distributions. E. L. LEHMANN and C. M. STEIN, University of California, Berkeley, California.

Critical regions are determined for testing a composite hypothesis, which are most powerful against a particular alternative among all critical regions whose probabilities under the hypothesis tested are bounded above by the level of significance. These problems have been considered by Neyman, Pearson and others, subject to the condition that the critical region be similar. In testing the hypothesis specifying the value of the variance of a normal

distribution with unknown mean against an alternative with larger variance, and in some other problems, the best similar region is also most powerful in the sense of this paper. However, in the analogous problem when the variance under the alternative hypothesis is less than that under the hypothesis tested, in the case of Student's hypothesis when the level of significance is less than $\frac{1}{2}$, and in some other cases, the best similar region is not most powerful in the sense of this paper. There exist most powerful tests which are quite good against certain alternatives in some cases where no proper similar region exists. These results indicate that in some practical cases the standard test is not best if the class of alternatives is sufficiently restricted

4. **On the Selection of Forecasting Formulas.** Paul G. Hoel, University of California, Los Angeles, California.

Given two competing formulas, $u = g(z_1, \cdots, z_m)$ and $v = h(z_1, \cdots, z_m)$, for forecasting a variable x, a significance test possessing optimum properties is designed for deciding whether one formula yields significantly better forecasts than the other The test, which turns out to be a Student t test, is constructed as a test of the hypothesis $H_0 : m_i = u_i$ against the alternative $H_1 : m_i = v_i$, $(i = 1, \cdots, n)$, in which it is assumed that the variables $x_1, \cdots, x_n$, corresponding to the n samples, are independently normally distributed with means m_i and variances $\sigma_i^2 = \sigma^2$.

5. **On the Power Function of the "Best" t-test Solution of the Behrens-Fisher Problem.** J. E. Walsh, Douglas Aircraft Company

The most powerful t-test solution of the Behrens-Fisher problem (one-sided and symmetrical) was obtained by Scheffé in *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 35-44. This note derives (approximately) the power efficiency of this t-test for the case in which the ratio of the variances of the normal populations is also known. Let the t-test be based on m sample values from the first normal population and n sample values from the second normal population, where $m \leq n$. For fixed values of m and n, a symmetrical t-test with significance level 2α has the same power efficiency as a one-sided t-test with significance level α. For one-sided t-tests with significance level α, the power efficiency is approximately $50[B + \sqrt{B^2 - 8(m+n)A}]/(m+n)$, where $B = 2 + (m+n)A + K_\alpha^2/2$, $A = 1 - K_\alpha^2/2(m-1)$, and K_α is the standardized normal deviate exceeded with probability α. This approximation is reasonably accurate for $m \geq 4$ if $\alpha = .05$, $m \geq 5$ if $\alpha = .025$, $m \geq 6$ if $\alpha = .01$, $m \geq 7$ if $\alpha = .005$ Intuitively the power efficiency of a test measures the percentage of available information per observation which is utilized by that test

6. **On Sequences of Experiments.** Charles Stein, University of California, Berkeley, California.

One performs a sequence of N experiments to decide between two simple hypotheses regarding probability distributions of certain observable quantities. At each stage there is a choice among L experiments and the one chosen yields a random variable. One wishes to achieve certain upper bounds α and β to the probabilities of first and second kind errors respectively, and, subject to these restrictions, to minimize the expected cost under a third hypothesis. The cost of each particular sequence of experiments is known. A solution is obtained, essentially by applying Lagrange's method and working back from the end of the experiment. This can be generalized to multiple decision problems. The results are applied to two-sample tests with the second sample of variable size, and to Wald's sequential analysis. As another problem, suppose (X_1, Y_1), $(X_2, Y_2) \cdots$ are independently distributed with bivariate normal distributions having mean ξ and covariance matrix Σ, both unknown. One tests $H_0 : \xi = 0$ against $H_1 : \xi'\Sigma^{-1}\xi = \delta$ A test (not necessarily optimum) valid within the usual approximation is obtained from the ratio of the p.d f.

of Hotelling's T^2 under H_1 to that under H_0. Analogous results hold for the multiple correlation coefficient, ratio of two variances and test for linear hypothesis.

7. The Effect of Selection Above Definite Lower Limits of Linear Functions of Normally Distributed Correlated Variables on the Means and Variances of Other Linear Functions. G. A. BAKER, University of California, Davis, California.

Sometimes certain variables in a system can be observed before other economically or socially important variables. These variables or linear combinations of them can be used as a basis of selection at given levels. The question is: How does selection on these earlier or more easily available variables affect the mean and variance of the economically or socially more important variables or, perhaps, linear functions of the more important variables. The general procedure is clear. We transform to a new system of variables which contains the linear functions on which selection is performed and the linear functions of which the means and variances are required as separate variables. The remaining new variables are eliminated by integration. The final calculation involves the numerical evaluation of integrals whose integrands are the product of polynomials and normal multivariate functions and whose limits depend on the given levels of selections. The general ideas are simple but the actual labor of computation in a given case is tedious. An example is considered in detail

8. An Inversion Formula for the Distribution of a Ratio of Random Variables. J. GURLAND, University of California, Berkeley, Calif.

The repeated Cauchy principal value of integrals applied to characteristic functions is used in obtaining inversion formulae for distribution functions. Let the random variables X_1 and X_2 have a joint distribution function with corresponding characteristic function $\phi(t_1, t_2)$. Suppose $P\{X_2 \leq 0\} = 0$. Let $\oint g(t)\,dt = \lim_{\substack{\epsilon \to 0 \\ T \to \infty}} \left(\int_{-T}^{-\epsilon} + \int_{\epsilon}^{T} \right) g(t)\,dt$ for any function $g(t)$. If $G(x)$ is the distribution function of X_1/X_2 then $G(x) + G(x - 0) = 1 - \frac{1}{\pi i} \oint \frac{\phi(t, -t_x)}{t}\,dt$. This formula is free of restrictions which accompany the formula given by Cramer in the case where X_1 and X_2 are independent; and differentiation extends a result of Geary to a much larger class of distribution functions. Further generalizations of the theory are obtained, and as an example the distribution function of the ratio of quadratic forms of random variables $X_1, X_2 \cdots X_n$ is considered in the case where $X_1, X_2 \cdots X_n$ have a multivariate normal distribution.

9. Independence of Parameters and Sufficient Statistics. E. W. BARANKIN, University of California, Berkeley, California.

The notions of *complete set of independent parameters* and *minimal set of sufficient statistics* are suitably defined for a class of families of probability densities $\{p(x_1, \cdots, x_n; \vartheta_1, \cdots, \vartheta_r)\}$, and the order of each of these sets is determined as the rank of a certain matrix. Second order continuous differentiability is eventually required of the function p; and certain other conditions are laid down, designed to ensure that the behavior of p in the large is similar to its behavior in the small when only continuous differentiability is assumed. The problem of determining the order of a minimal set of sufficient statistics is made, by certain device, to become identical in character with that of finding the order of a complete set of independent parameters. (This is in the nature of these concepts.)

An explicit method is given for finding a complete set of independent parameters and a minimal set of sufficient statistics.

(Presented December 30, 1947 at New York at the Annual Meeting of the Institute)

1. **Distribution of the Circular Serial Correlation Coefficient for Residuals from a Fitted Fourier Series** (*Preliminary Report*). R. L. ANDERSON, University of North Carolina, Raleigh, North Carolina and T. W. ANDERSON, Columbia University.

Given a set of N observations $\{X_i\}$, which are defined as follows:

$$X_i - \mu_i = \rho \cdot (X_i - L - \mu_i - L) + \epsilon_i ,$$

where the residuals $\{\epsilon_i\}$ are assumed to be normally and independently distributed with zero means and equal variances and L is the lag. A statistic for testing the null hypothesis: $\rho = 0$ is ${}_LR$, the circular serial correlation coefficient of residuals e_i from a regression line fitted by least squares: $X_i = M_i + e_i$. The following regression line is considered:

$$M_i = a_0 + \sum_k{}' a_k \operatorname{Cos} \frac{2\pi ki}{N} + \sum_k{}' b_k \operatorname{Sin} \frac{2\pi ki}{N},$$

where k ranges over some subset of the integers $1, 2, \cdots, \frac{1}{2}(N-1)$ or $\frac{1}{2}(N)$, depending on whether N is odd or even (if N is even, $b_{\frac{1}{2}N}$ is not used). Hence ${}_LR$ is defined as:

$${}_LR = \frac{e_1 e_{L+1} + e_2 e_{L+2} + \cdots + e_N e_{L+N}}{\Sigma e_i^2},$$

with $e_{i+N} = e_i$.

The distribution of this ${}_LR$ has the same general form as that presented by R. L. Anderson for $\rho = 0$ ["Distribution of the serial correlation coefficient," Annals of Math. Statistics 13:1-13(1942)]; and for $\rho \neq 0$ by W. G. Madow ["Note on the distribution of the serial correlation coefficient," *Annals of Math. Statistics* 16 308-310(1945)].

For M_i consisting of terms of only *one* period, $\frac{N}{k} = 2, 3, 4, 6, 12$ and 24, exact values of the 1% and 5% significance levels of ${}_1R$ have been computed for $N = 12$ and 24. Approximate significance levels have been computed for $N = 12(12)96$. More of the exact significance levels are being computed, and all computations will be extended to include some multiple periods and some lags greater than 1.

2. **Some New Methods for Distributions of Quadratic Forms.** HAROLD HOTELLING, Institute of Statistics, University of North Carolina, Chapel Hill.

Any homogeneous quadratic form in normally distributed variates of zero means has the same distribution as $q = \frac{1}{2}(a_1x_1^2 + \cdots + a_nx_n^2)$, where the a_i are roots of a determinantal equation based on the coefficients of the given form and the parameters of the normal distribution, and where the x_i are normally and independently distributed with zero means and unit variances. We take $\Sigma a_i = n$, and begin by expanding the distribution of a positive definite form in a series of powers of q whose coefficients are polynomials in the reciprocals of the a_i. This series shows the analyticity of the function, which is then expressed as the product of a χ^2 distribution function of a series of Laguerre polynomials with coefficients which are simple polynomials in the moments of the a_i. Indefinite forms and certain ratios of forms are dealt with by convolutions of these series and by other means.

3. **Frequency Functions Defined by the Pearson Difference Equation.** LEO KATZ, Michigan State College, East Lansing, Michigan.

Frequency "links" formed from the Pearson difference equation provide an efficient means of fitting functions to observed distributions. These links, involving three constants which are determined by the first four moments of the observed series, correspond to a three-parameter family of discrete frequency functions. This family of functions is just as broad as that defined by the differential equation, containing functions of equally diverse types; in addition, it has the very important advantage that the graduation process is the same for any type. Further, the simpler functions of the family all correspond to points lying in one plane of the parameter space. This plane, giving a two-parameter family of functions (depending upon the first three moments), is studied intensively, rather complete results being obtainable for areas, moments, sampling characteristics of moments, etc. It is also shown that the problem of discrimination among simple discrete frequency functions for graduating observed data is resolvable (in the plane) to the sampling distribution of one statistic. A special case of the two-parameter family depending on only the first two moments was previously discussed

4 **Distribution of the Sum of Roots of a Determinantal Equation under a Certain Condition.** D. N. NANDA, Institute of Statistics, University of North Carolina, Chapel Hill

Let $x = \| x_{ij} \|$ and $x^* = \| x^*_{ij} \|$ be two p-variate sample matrices with n_1 and n_2 degrees of freedom Then $S = xx'/n_1$ and $S^* = x^*x^{*\prime}/n_2$ are, under the null hypothesis, independent estimates of the same population covariance matrix. The distribution of a root, specified by its rank order, of the determinantal equation $| A - \theta(A + B) | = 0$, where $A = n_1S$ and $B = n_2S^*$, has already been given by S. N. Roy, and by the author, who has also obtained the limiting distribution of any root when one of the samples becomes infinitely large The moment generating function of the sum of the roots when $n_1 = p \pm 1$ can be derived from the limiting distribution of the largest root. The probability distributions of the sum of roots under this condition have been formulated for the determinantal equations having two, three, and four roots. The moments of these distributions have also been obtained The method is applicable for the determinantal equation of any order. These probability distributions can easily be tabulated, as they involve only simple algebraic and incomplete beta functions.

5. **Applications of Carnap's Probability Theory to Statistical Inference.** GERHARD TINTNER, Iowa State College, Ames, Iowa.

The new theory of probability of Rudolf Carnap ("On inductive logic," *Philosophy of Science*, vol 12, 1945, pp 72 ff "The two concepts of probability," *Philosophy and Phenomenological Research*, vol. 5, 1944, pp 513 ff.) introduces a distinction between probability$_1$, the degree of confirmation, and probability$_2$, related to relative frequency. It is believed, that the ideas developed are useful in clarifying the problems of statistical inference.

As an example, consider the case of "inverse inference," i.e. inference from a sample to the population. The evidence is that in a sample of size s there are s_1 individuals with a certain property M and $s_2 = s - s_1$ without the property. The hypothesis is that in the population consisting of n individuals there are n_1 individuals with property M and $n_2 = n - n_1$ individuals without this property. The degree of confirmation is then:

$$c^* = \frac{\binom{n_1 + w_1 - 1}{s_1 + w_1 - 1}\binom{n_2 + w_2 - 1}{s_2 + w_2 - 1}}{\binom{n + k - 2}{n - s}}$$

In this formula we have: w_1 the logical width of the property M, w_2 the logical width of the property non-M, $k = w_1 + w_2$. It should be noted that for $w_1 = w_2 = 1$ the formula becomes the classical result, i.e. a term of the hypergeometric distribution.

This idea may be applied to statistical estimation. We could for instance choose n_1 in such a fashion that c^* becomes a maximum. This would be estimation by the principle of maximum degree of confirmation, analogous to maximum likelihood. In a similar fashion we may also use c^* to establish limits for n_1 similar to confidence or fiducial intervals.

6. **Circular Probable Error of an Elliptical Gaussian Distribution.** HALLETT H. GERMOND, S. W. Marshall & Co., Consulting Engineers, Washington, D. C.

Preliminary tables are presented, giving the radii of distribution-centered circular cylinders enclosing various percentages of the volume under an elliptical bivariate Gaussian surface. These tables are further interpreted in terms of a correlated bivariate Gaussian distribution. The application of these tables to impact analysis is illustrated.

(Presented December 29, 1947 at the Chicago Meeting of the Institute)

1. **The Asymptotic Analogue of the Theorem of Cramér and Rao.** HERMAN RUBIN, Institute for Advanced Study, Princeton, N. J.

The author generalizes the results of Cramér and Rao on the minimum variance of estimates to the case of the asymptotic distribution of an estimate. He shows that if certain regularity conditions are satisfied, the formula given by Cramér and Rao remains valid. The main results are obtained in the case of consistent estimates, but with a stronger set of hypotheses, the results remain true for estimates which are not consistent. The method used to obtain these results is to construct statistics to which the theorem of Cramér and Rao can be applied, and whose variance converges to the variance of the limiting distribution. This procedure is also applied to the case in which there is no limiting distribution, and in which two sequences of distributions are considered which act as if they approach each other.

BOOK REVIEWS

Sequential Analysis *Abraham Wald.* John Wiley and Sons, Inc. pp. vi, 212, $4.00.

REVIEWED BY M. A. GIRSHICK

Douglas Aircraft Company

The development of sequential analysis as a new tool of statistics is by and large the work of Abraham Wald. This fact in itself would make the appearance of a book by him on this subject an important event. However, Wald in this book did more than discuss the present status of sequential theory. He has, in fact, written a very lucid treatise on the general subject of statistical inference—a treatise which is likely to have great influence on statistical thinking.

While this book is not written for the mathematically untrained, a knowledge of differential and integral calculus will suffice to follow all the arguments except perhaps for some sections in the appendix where the more complicated proofs have been placed.

The main body of this book is divided into 3 parts and 11 chapters. Part I, covering chapters 1 to 4 inclusive, deals with the general theory of the sequential probability ratio test. Chapter 1 introduces in an elementary fashion the notion of probability distributions, tests of hypotheses and the Neyman-Pearson theory of two-valued decisions based on a fixed sample size. In Chapter 2, the general notion of a sequential test procedure is introduced and the operating characteristics of such tests are discussed. Chapter 3 deals with the sequential probability ratio test for testing a single hypothesis against a single alternative. Here the boundaries of this sequential criterion are expressed in terms of the risks, the operating characteristic and the average sample number functions are developed and bounds are obtained for the errors arising from truncation and neglect of excess over the boundaries. Chapter 4 presents a sequential theory for testing simple and composite hypotheses against a set of alternatives. The fundamental idea introduced is the concept of a weight function in the parameter space which permits handling composite hypotheses, or simple hypotheses with many alternatives, by means of the sequential probability ratio test.

Part II of this book, consisting of chapters 5 to 9 inclusive, deals with the applications of sequential analysis to special problems. Chapter 5 contains a discussion of the binomial case with specific reference to lot-by-lot acceptance inspection. Of special interest in this chapter is the derivation of the exact characteristic function for a large class of tests and the development of upper and lower limits for the effect of grouping on the OC and ASN curves. Chapter 6 deals with the problem of double dichotomies. A procedure for testing the difference between the parameters of two binomial distributions is developed

for the fixed size as well as the sequential procedure. Chapters 7, 8, and 9 are concerned with the application of sequential analysis to the normal distribution. In these chapters the sequential probability ratio test is applied to hypotheses concerning the mean of a normal distribution when the variance is known, when the variance is not known (non-central t case) and hypotheses concerning the variance when the mean is known and when the mean is not known.

Part III consists of two short chapters and deals with multi-valued decisions and sequential interval estimation. The results in these chapters are not definitive answers to the two outstanding problems in statistical inference but are merely suggestive of a possible approach to them. Nevertheless, from the point of view of stimulating future research these 2 chapters are perhaps the most valuable sections of this book. The reader, having been exposed in the previous chapters to various tests the outcome of which is a two-valued decision, is naturally led in Chapter 10 to the consideration of tests the outcome of which is a multi-valued decision. The notion of a risk function, introduced elsewhere by the author in the non-sequential case, is again used as the main tool in handling multi-valued decisions sequentially. In Chapter 11 the important problem of setting up confidence intervals of fixed length by means of a sequential procedure is discussed and a possible method for accomplishing this is indicated.

As was previously noted, the main theorems on sequential analysis are contained in the Appendix and since they have all been previously published in the Annals they will not be mentioned in the present review. The Appendix, together with the main body of the book form a fairly exhaustive treatment of sequential theory. A notable exception to this is the lack of any mention of the published research on sequential point estimation. This is probably accounted for by the fact that this research came too late to be included in the book. Other minor omissions that may be noted are references to the generalization of the Fundamental Identity to more than one dimension and other theorems on sequences of functions of random vectors which have appeared in print. Also no mention is made of the similarity of sequential analysis to the problems of the random walk and the gambler's ruin. This, in the opinion of the reviewer, is regrettable.

This book will make a very suitable companion to the book *Sequential Analysis of Statistical Data: Applications* prepared by the Statistical Research Group, Columbia University (see review by J. W. Tukey, *Ann. of Math. Stat. Vol.* xviii, 1947). While there is some overlap in the material covered, the two books differ in emphasis. Wald's book, though not highly technical, is more in the nature of a textbook on the theory and application of sequential analysis. The SRG book on the other hand, was prepared mainly for statisticians who may wish to use sequential analysis in practice. The latter book is therefore more detailed and puts less emphasis on the theoretical aspects of the sequential procedure.

The book is surprisingly free of typographical errors which is a tribute to the high quality of the editorship.

Statistical Methods. *George W. Snedecor.* Ames, Iowa: The Iowa State College Press, Inc., 1946; pp. xvi, 485. $4.50.

REVIEWED BY FREDERICK MOSTELLER

Harvard University

Statistical Methods is a non-mathematical treatment of modern experimental statistics. Few non-mathematical books are available that treat such topics as confidence limits, use of transformations, and analysis of variance and covariance in the detail presented by Snedecor. The examples are largely, but not entirely, drawn from agriculture and animal husbandry. The exercises for students are extensive and thought-provoking.

Unlike most non-mathematical texts the book under review does not spend pages and pages on methods of recording frequencies and methods of computing countless moments which are seldom used in the later developments of the text. There is no long exasperating discussion of kurtosis and skewness; and there is no parade of qualitative Greek names for categorizing frequency distributions.

The reviewer has used this book for teaching a second course in statistics to social science majors with reasonable success. The main disadvantage was the biological nature of most of the examples, but until some author writes a comparable book using social science examples, the reviewer will continue to use Snedecor's material for a large part of the course.

The main differences between the Third and Fourth Editions of this text have been adequately summarized by Snedecor:

"(i) greater emphasis has been placed on the theoretical conditions in which the various statistical methods have validity, and concurrently (ii) on the conduct of the experiment so as to incorporate in the data the information desired; (iii) estimates and fiducial statements have been brought into equal prominence with tests of hypotheses; (iv) there is increased reliance on experimental samplings to exemplify distribution theory; (v) the treatment of correlation and of experimental designs has been expanded; and (vi) the methods for disproportionate subclass numbers have been extended to include all those necessary for ordinary needs." Some more obvious changes in the Fourth Edition are the entirely new type and summaries which are included at the end of some of the chapters. The practice of using random sampling numbers (iv) to help explain theory has long been employed by teachers of statistics, but few authors have taken as much advantage of this technique as has Snedecor. In the Fourth Edition confidence intervals are widely used (iii). The author uses the adjectives "confidence" and "fiducial" more or less interchangeably, but it is the reviewer's opinion that it is the Neyman concept rather than the Fisherian that predominates. It should be remarked that this is one of the few texts that give the students the idea that in linear regression we do not predict y with the same accuracy for every x even when linearity and homoscedasticity hold (v).

The main emphasis of the book is on the analysis of variance. The author succeeds extremely well in showing the student how to carry out the analysis

even at rather complex levels. On some other points he was not quite so successful. For example, the reviewer feels that the meaning of "interaction" was never gotten across, and that for the student the higher order interactions are still just things to be computed. Furthermore in attempting to make sure that the student understands how to do the computation the author often does not encourage the student to take any overall view of the data before blindly starting to compute. In addition, reasons for doing the experiment are sometimes vague and the conclusions are often couched only in the jargon of analysis of variance. Therefore, the student seldom gets an opportunity to find out what kinds of recommendations might reasonably be made as the result of an experiment. Perhaps the worst example is on pages 275–280 Here the experiment deals with yield of wheat in 48 pots, with two series of soil treatments, humus and chemical. Anyone glancing over the results of the experiment will be startled to find that every yield from pots with "no humus treatment" (12 observations) is greater than any yield with "humus treatment" (36 observations). The reader will be further startled to find that all the evidence tends to support the notion that "no chemical treatment" is at least as fruitful as any of the chemical treatments tried. However, Snedecor says "The striking feature of this experiment is the discrepance among the subclasses The chemicals applied to one humus treatment produced yields out of accord with those from other humus treatments." Snedecor then pushes on to a more subtle analysis. The reviewer feels that here as elsewhere in the book the author occasionally forgets that the extended analysis looks rather ridiculous unless the practicality of applying the technique is discussed. The example considered here is one in which the point could profitably be made that everyone can see from a visual examination of the data what the results of the experiment show. The analysis backs up the student's common sense appraisal of the situation and gives him more confidence in and understanding of the method when it is applied in more delicate situations. It seems to the reviewer that too many times the application of the analysis of variance obfuscates the main point of the experiment. In the haste to get to the computations and the comparisons of interactions and errors the author frequently neglects to impress the student with the fundamental differences between means and their ultimate interpretation. However, the author does bring out clearly the notion of the various estimates of variance, a subject frequently neglected.

In the next to last chapter the binomial and Poisson distributions are discussed. In this connection the inverse sine and the square root transformations are treated briefly, as is the logarithmic transformation It is surprising that no indication is given of the theoretical variances when the inverse sine and square root transformations are used. The theoretical discussion of the transformation is limited to the remark that these transformations tend to make the variance independent of the means, but there is no indication of the further advantages. This is surprising because in a much earlier chapter the use of Fisher's transformation for correlation coefficients was treated quite adequately. It seems

to the reviewer that in a later edition the use of transformation might well be moved forward in the book, and that the theoretical and practical implications might be treated more thoroughly.

As in most other texts the final chapter "Design and Analysis of Samplings" needs very considerable expansion.

The book begins (Chapter 1) with a consideration of the sampling of attributes, inferences that can be drawn about the population, confidence limits, use of chi-square in a 1 x 2 table, and some discussion of the use of ratios, rates, and percentages. Measurement data is then (Chapter 2) discussed including the computation and application of the mean, range, standard deviation, probable deviation, median, and quartiles. The concepts of null hypothesis and confidence limits are introduced in Chapter 2 and elaborated in Chapter 3 which concerns sampling from a normally distributed population, random samples, distribution of the mean, variance, standard deviation, and of t. The comparison of two groups in contrast to individuals is treated in Chapter 4 including groups with different numbers of individuals. Chapter 5 provides material on short cut methods of computation using calculating machines, code numbers are explained, suggestions about significant numbers and rates and percentages are given, and the use of the ratio range/sigma is introduced.

After considering linear regression and correlation (Chapters 6, 7) the author relates the two notions, and then goes on to consider some interesting special cases of correlation. Chapter 8 deals with large sample methods. Chapter 9 concerns enumeration data with more than one degree of freedom, discusses adjustments of chi-square and its computation with large numbers of degrees of freedom, and describes the analysis of 2 x 2 x 2, R x 2, and R x C tables. The computation of the analysis of variance for two or more groups of measurement data and with two or more criteria of classification: variance ratio F, use of Latin square, analysis with disproportionate subclass numbers, and the use of randomized blocks are considered in Chapter 10 and 11, while analysis of covariance is treated in Chapter 12 (22 pages). Multiple regression including partial and multiple correlation coefficients, tests of significance and confidence limits are handled in Chapter 13 and curvilinear regression considered in Chapter 15. Chapter 16 deals with binomial and Poisson data, and Chapter 17 discusses the design and analysis of sampling, including sampling from a homogeneous or small population and the effectiveness of stratification.

It seems to the reviewer that at the present time one would be hard put to find a better statistics text written at this level.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Franz L. Alt, who has been with the Econometric Institute, New York, as Assistant Director of Research, is now Deputy Chief of the Computing Laboratory at the Ballistic Research Laboratories, Aberdeen Proving Ground, Aberdeen, Maryland.

Mr A. George Carlton has accepted a position as Assistant Professor of Mathematics at the University of Illinois.

Assistant Professor Paul R. Halmos, University of Chicago, Chicago, Illinois is on leave for the academic year. He is spending the year at the Institute for Advanced Study, Princeton, New Jersey on a Guggenheim Fellowship and will return to the University of Chicago in September, 1948.

Mr. Henry F. Hebley of the Pittsburgh Coal Co. spent most of last summer in Eastern Europe carrying out a survey on coal production and fuel availability in Poland This work was carried out in the interest of the International Bank for Reconstruction and Development.

Dr. Harold D. Larsen, former Associate Professor at the University of New Mexico, has joined the faculty of Albion College, Albion, Michigan.

Mr Dickson H. Leavens has resigned as Research Associate of the Cowles Commission for Research in Economics. He will continue as Managing Editor of *Econometrica* and may be addressed at 1632 Wood Avenue, Colorado Springs, Colorado.

Professor S. B. Littauer, who has been Chairman of the Mathematics Department, Newark College of Engineering, Newark, New Jersey, has now accepted an associate professorship in the Department of Industrial Engineering, Columbia University.

Professor Harris F. MacNeish, who has been Chairman of the Department of Mathematics at Brooklyn College since its foundation in 1930, has resigned to accept a visiting professorship in Mathematics at the University of Miami, Coral Gables, Florida

Mr. Clifford J. Maloney has resigned a position as Research Associate in the Statistical Laboratory of Iowa State College to serve as Chief, Statistics Branch, Camp Detrick, Frederick, Maryland, an agency of the Chemical Corps of the United States Army.

Mr. Monroe L. Norden, who has formerly been with the Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland, has accepted a research position in theoretical or mathematical statistics at the Douglas Aircraft Co., Santa Monica, California.

Mr. W. E. Pattee has resigned his position as statistical engineer with the Canadian Industries Limited, Skawinigan Falls, Quebec and has accepted a position as senior chemist, Ottawa Mill, E. B Eddy Company, Hull, Quebec.

Mr. Robert I. Piper, who was formerly plant staff assistant at the Southern California Telephone Company of Los Angeles, has been transferred to the systems office of the Pacific Telephone and Telegraph Company. He will assist in planning and analysing sampling surveys of the wages rates prevailing in the Pacific coast states in which the company operates.

Mr. Herbert Solomon, who was formerly an instructor at the College of the City of New York, has accepted an assistant professorship in the Mathematics Department, Newark College of Engineering, Newark 2, New Jersey.

Dr. A G. Swanson, formerly an assistant chairman of the Department of Mathematics and Mechanics at the General Motors Institute, Flint, Michigan, has accepted an associate professorship in the Department of Mathematics, Gustavus Adolphus College, St. Peter, Minnesota.

A federal center of applied mathematics—the National Applied Mathematics Laboratories—has been established as a division of the National Bureau of Standards. The new organization is oriented around modern mathematical statistics as applied to the physical and engineering sciences and to the development and use of modern high speed computing. The applied mathematics laboratories include four separate laboratories: the Institute of Numerical Analysis, the Computation Laboratory; the Statistical Engineering Laboratory; and the Machine Development Laboratory.

Two members of the Institute have been given important positions in this organization. Dr. John Curtiss, who has been Director's Assistant in Applied Mathematics at the Bureau of Standards, has been named Chief of the National Applied Mathematics Laboratories Dr. Churchill Eisenhart has been appointed head of the Statistical Engineering Laboratory.

Statistical Summer Sessions at the University of California, Berkeley

Following the encouraging experience of last year the University of California offers statistical programs in the two Summer Sessions of 1948. The teaching staff is as follows:

Raj Chandra Bose, Professor of the University of Calcutta, India.

Miss Evelyn Fix, Lecturer at the University of California, Berkeley.

Erich L. Lehmann, Assistant Professor of the University of California, Berkeley.

Michel Loève, Reader at the University of London, England.

Jerzy Neyman, Professor of the University of California, Berkeley.

Abraham Wald, Professor of Columbia University, New York.

Courses in statistics are offered on both the graduate and the undergraduate levels. The graduate courses, all given during the First Summer Session, June 21

to July 31, are meant primarily for students who either have already obtained their Ph.D. degree or are working towards it. Therefore, apart from formal classes, it is proposed to hold extensive seminars in which the work of students will be discussed. No specific prerequisites to graduate courses will be required. However, to benefit from the courses, the students must be generally familiar with the theory of statistics. In addition, course 272 and especially 271 will require a reasonable knowledge of the theory of functions.

There will be two undergraduate courses offered, course S12 during the First Summer Session, June 21 to July 31, and course S113 during the Second Summer Session, August 2 to September 11. Both of these courses were recently introduced into the curriculum and are prerequisites to more advanced courses in statistics They are offered during the Summer Sessions for the benefit of students, otherwise advanced, who plan to attend more advanced courses in statistics during the fall semester. Besides, course S12 is recommended for students who do not intend to specialize in statistics but wish to acquire some knowledge of this subject as a part of their general education.

The Statistical Laboratory will be available for students doing research.

First Summer Session

S12 Elements of Probability and Statistics	Mr. Lehmann
271. Random Functions	Mr. Loève
272. Sequential Analysis	Mr. Wald
273. Design of Experiments	Mr. Bose
S290s. Seminar in Theory of Statistics	Mr. Loève, Mr Wald
290t Seminar in Design of Experiments.	Mr Bose
S295. Individual Research.	Mr. Bose, Mr Loève, Mr. Neyman, Mr. Wald

Second Summer Session

S113. Second Course in Probability and Statistics.	Miss Fix.

Statistical Sessions at Alabama Polytechnic Institute

Professor George W. Snedecor, President of the American Statistical Association and Research Professor of Statistics at Iowa State College, will be Visiting Research Professor of Statistics at Alabama Polytechnic Institute during the Spring Quarter, from March 22 to June 4, 1948. Professor Snedecor will lecture on Statistical Experimental Design and will be available for statistical consultations.

The newly formed Stastistical Laboratory at A.P I. will also offer a course in Survey Sampling during the Spring Quarter to be taught by the Director, Professor T. A. Bancroft. Conferences in applied statistics for research workers in the lower southeastern states are being scheduled during the time of Professor Snedecor's visit.

New Members

The following persons have been elected to membership in the Institute

(September 1 to November 30, 1947)

Afzal, M., M A. (Panjab, India) Graduate student at Columbia Univ., *1038 John Jay Hall, Columbia University, New York 27, New York.*

Billeter, Ernest P., Ph.D. (Univ. of Basle) Scientific Assistant (Statistical Office, Zurich) *Turnerstrasse 23, Basle, Switzerland.*

Bishop, David James, M Sc. (London) Head of Operational Research Section of British Iron and Steel Research Association, 11 Park Lane, London W. 1., England.

Brooks, Hamilton, B.Sce (Univ. of Pittsburgh) Design Engineer, Westinghouse Electric Corp., *P.O. Box 983 E. Pittsburgh, Pennsylvania.*

Craw, Alexander R., M S. (Univ. of Notre Dame) Instructor in Math., U. S. Naval Academy, Annapolis, Maryland.

Edwards, Daisy M., A M. (Columbia Univ.) Lecturer in Statistics, University of London, Institute of Education, *1, Oakfield Court, Queens Road, Weybridge, Surrey, England.*

Havermark, K. Gunnar, Chief of Division, Royal Social Board, *Lagerlofsg 8, Stockholm, Sweden.*

Hollingsworth, Charles A., Ph.D., (State Univ. of Iowa) Research Chemist, *604 Maple Ave., Waynesboro, Virginia.*

Hurd, Cuthbert C., Ph.D. (Univ. of Ill.) Plant Statistician, Carbide and Carbon Chemicals Corp., Oak Ridge, Tenn.

Isaacson, Stanley L., M.A. (Johns Hopkins Univ.) Graduate student at Columbia Univ , *2523 Loyola Southway, Baltimore, Maryland*

May, Kenneth, Ph.D., (Univ of Calif.) Assistant Professor of Mathematics, Carleton College, Northfield, Minnesota.

Mirsky, Robert, A M. (Johns Hopkins Univ.) Graduate student at Columbia Univ., *7 West 705th Street, Shanks Village, Orangeburg, New York.*

Mulhall, Harold, B.Sc (Sydney) Lecturer in Mathematics, Department of Mathematics, University of Sydney, Australia.

Palm, Conny, Ph D. (Stockholm) Docent, *Ynglingar 11, Djursholm, Sweden.*

Pease, Katharine, A.M. (Smith College) Instructor in Psychology, Barnard College, Columbia University, New York 27, New York.

Peckham, Cyril G., M.S. (Univ. of Ill.) Assistant Professor of Mathematics, University of Dayton, Dayton 9, Ohio

Peterson, Raymond P., Jr., B.A. (Univ. of Calif., Los Angeles) Assistant in Mathematics, University of California, Los Angeles, Calif., *10729 Ashton Ave., Los Angeles 24, California*

Pike, Eugene W., Ph.D., (Princeton) Member McFarlan, Groth & Pike, *510 Audubon Ave., New York 33, New York.*

Pitman, Edwin J. G., M.A. (Univ. of Melbourne) Professor of Mathematics, Univ. of Tasmania, Hobart, Tasmania.

Rigby, Fred D., Ph D., (Univ. of Iowa) Mathematician, Office of Naval Research, *P.O. Box 234, Falls Church, Virginia.*

Smith, Clarence DeWitt, Ph.D. (Univ. of Iowa) Associate Professor of Statistics, Box 2686, University, Alabama.

Srinivasan, T. K., M A (Madras) Assistant Lecturer, Mathematics Department, Raja's College, Pudukkottah, S-I-R, South India.

Straubel, Morgan P., Quality Control Analyst, *4124 Ivanrest Road, Grandville, Michigan.*

Taylor, William F., A.B (Univ. of Calif., Berkeley) Associate, School of Public Health, *3042 Wheeler St , Berkeley, California.*

Trindade, Mario, Chief of the Statistical Division of the Instituto de Resseguros do Brazil, *Rua Senador Soares 33, ap. 201, Rio de Janeiro, Brazil.*

Von Schelling, Hermann, Ph.D (Univ of Berlin) Naval Medical Research Laboratory, U. S. Submarine Base, New London, Conn.

Whidden, Phillips, A.B. (Harvard) Part-time Instructor in Mathematics, Carnegie Institute of Technology, Pittsburgh 13, Pa.

Wolman, William, B.B A. (College of City of New York) Statistician, New York State Division of Housing, *295 Parkside Avenue, Brooklyn 26, New York.*

Woodbury, Lowell A., Ph.D. (Univ of Michigan) Assistant Professor of Physiology, Dept of Physiology, University of Utah Medical School, Salt Lake City 1, Utah.

Yusuf, Mohammad, M.A. (Aligarh Muslim Univ , India) Graduate student at Columbia University, *208, Furnald Hall, Columbia University, New York 27, New York*

REPORT ON THE BERKELEY MEETING OF THE INSTITUTE

The thirtieth meeting of the Institute of Mathematical Statistics was held in Berkeley, California on Monday and Tuesday, December 22 and 23, 1947. The meeting was attended by approximately 70 persons, including the following 31 members of the Institute:

G. A. Baker, G. G. Beckstead, B. M. Bennett, R. U. Bonner, Frances L. Campbell, E. L. Crow, Dorothy Cruden, W. J Dixon, R. Dorfman, G. G. Eldredge, E. A. Fay, Evelyn Fix, M A Girshick, J. Gurland, T. E. Harris, W. L. Hart, J. L Hodges, Jr., P. G Hoel, H. M. Hughes, T. A. Jeeves, H. S Konijn, G M. Kuznets, E L. Lehmann, R. B. Leipnik, J. Neyman, Gladys Rappaport, H Scheffé, T. W. Simpson, C M. Stein, J. E. Walsh and H Working.

The Monday morning program, with Professor J. Neyman presiding, consisted of the following contributed papers:

1. *The Performance Characteristic of Certain Methods for Obtaining Confidence Intervals.* Mr B M Bennett, University of California, Berkeley.
2. *Some Further Results on the Bernoulli Process* Dr T. E Harris, Douglas Aircraft Company.
3. *Most Powerful Tests of Composite Hypotheses I. Normal Distributions.* Dr. E L. Lehmann and Dr. C. M. Stein, University of California, Berkeley.
4. *On the Selection of Forecasting Formulas.* Professor P. G. Hoel, University of California, Los Angeles.

The Monday afternoon program, with Professor H. Scheffé presiding, also consisted of contributed papers as follows:

1. *On the Power Function of the "Best" t-test Solution of the Behrens-Fisher Problem.* Dr J E Walsh, Douglas Aircraft Company.
2. *On Sequences of Experiments* Dr C. M Stein, University of California, Berkeley.
3. *The Effect of Selection above Definite Lower Limits of Linear Functions of Normally Distributed Correlated Variables on the Means and Variances of Other Linear Functions.* Professor G A. Baker, University of California, Davis.
4. *An Inversion Formula for the Distribution of a Ratio of Random Variables.* Dr. J. Gurland, University of California, Berkeley.
5. *Independence of Parameters and Sufficient Statistics.* Dr E W Barankin, University of California, Berkeley.

The Tuesday morning session, with Professor R. A. Gordon presiding, was devoted to the following invited and contributed papers on econometrics:

1. *Remarks on the Theory of Indices* Professor G C. Evans, University of California, Berkeley.
2. *Interrelations of Theory and Statistical Research in Economics.* Professor H Working, Stanford University
3. *Statistical and Case Methods in a Study of Labor Mobility.* Professor D. McEntire, University of California, Berkeley. Discussion · Dr M Lipton, University of California, Berkeley.

4 *Distributions Associated with Continuous Stochastic Processes*
Dr R B Leipnik, University of California, Berkeley
5 *On Some Methods of Evaluating Railway Costs* (By title)
Miss Evelyn Fix, University of California, Berkeley

There was a dinner on Monday evening for members and guests at the Hotel Claremont and an informal discussion and coffee on Tuesday afternoon.

REPORT ON THE NEW YORK MEETING OF THE INSTITUTE

The Tenth Annual Meeting of the Institute of Mathematical Statistics was held at the Commodore Hotel, New York City, on December 28-30, 1947. The meeting was held in conjunction with the American Statistical Association. The following 173 members of the Institute were in attendance:

F S Acton, R L Anderson, H E Arnold, L A Aroian, M. Astrachan, H M Baldwin, W D. Baten, R. E. Bechhofer, G W Beebe, M H. Belz, A A Bennett, A. J. Berman, A. Blake, C I Bliss, P Boschan, A H Bowker, A. E Brandt, T. H Brown, M A Brumbaugh, M C. Bruyere, P T Bruyere, T A Budne, R. W Burgess, R. S. Burington, B H Camp, G C. Campbell, P G. Carlson, Jr , U. Chand, H. Chernoff, Kai-Lai Chung, P C. Clifford, W. G. Cochran, D. D Cody, J Cornfield, G. M. Cox, J. H Curtiss, J. F. Daly, G B. Dantzig, D. G. Deihl, H F. Dorn, A J Duncan, C W Dunnett, D. Durand, J Dutka, P S. Dwyer, G L. Edgett, C Eisenhart, B. Epstein, M. W Eudey, W D Evans, Will Feller, C D. Ferris, C. B. Fine, M. M Flood, L R Frankel, J E. Freund, B. Friedman, Hilda Geiringer, M. A Geisler, H H Germond, M A Girshick, Abraham Golub, C.H. Graves, S W Greenhouse, J A. Greenwood, T N E Greville, J. I Griffin, E T Gumbel, M. Gurney, K W. Halbert, Max Halperin, M H. Hansen, T. E Harris, B Harshbarger, Alex Hart, P. M Hauser, J D Heide, L H Herbach, M W Hirsch, Harold Hotelling, H M Humes, C. C. Hurd, S Jablon, C. M Jaeger, A. S Kaitz, Leo Katz, T. L Kelley, L S Kellogg, L. F Knudsen, A. K. Kury, Jack Laderman, M LeLeika, Joseph Lev, Howard Levene, J E Lieberman, Julius Lieblem, S. B Littauer, Eugene Lukacs, Geo. A Lundberg, J. C. McPherson, Benjamin Malzberg, Sophie Marcuse, E S Marks, H C Mathisen, J. W. Mauchly, A. L. Mayerson, Margaret Merrell, E. B Mode, E C Molina, M. E. Moore, D. J. Morrow, J. E Morton, Jack Moshman, Hugo Muench, D N. Nanda, M G Natrella, Doris Newman, G. E Nicholson, Jr , Harold Nisselson, Nilan Norris, H. W. Norton, P S. Olmstead, A L. O'Toole, A E. Paull, C. N Payne, Katherine Pease, M P. Peisakoff, E. W Pike, O A. Pope, G B. Price, L J Reed, J S Rhodes, S F Robinson, A C Rosander, Ernest Rubin, P. J. Rulon, Rose Sachs, Frank Saidel, Arthur Sard, M M. Sandomire, F. E Satterthwaite, E D. Schell, Bernice Scherl, O. N Serbein, R. G. Seth, Harry Shulman, Rosedith Sitgreaves, C. DeW Smith, G. W. Snedecor, Herbert Solomon, D. E South, Arthur Stein, G. T Steinberg, Joseph Steinborg, A. I Sternhell, S. A. Stouffer, J V. Sturtevant, B. R. Suydam, W R. Thompson, Gerhard Tintner, J W. Tukey, D. F Votaw, Jr., A. J. Wadman, H. M. Walker, Dzung-shu Wei, Sidney Weiner, Samuel Weiss, Sophie R. Wilkey, R. I. Wilkinson, S. S. Wilks, C. P. Winsor, Jacob Wolfowitz, W. J. Youden.

The first session, a joint session with the American Statistical Society, was held on the morning of December 28 and was devoted to the topic *The Teaching of Statistics*. Professor W G Cochran of North Carolina State College presided. A paper entitled *Three Recent Reports Dealing with the Teaching of Statistics*,

the Training of Statisticians and the Crisis in Statistical Personnel was presented by Dr. James D. Paris of the Metropolitan Life Insurance Company. Many members participated in the general discussion which followed.

The second session on *The Teaching of Statistics* also with the American Statistical Association, was held at 1:15 P.M. Professor Francis G. Cornell of the University of Illinois was chairman. The main paper of the session was the paper by Professor George W. Snedecor of Iowa State College entitled *Syllabus for a Proposed Course in Basic Statistics*. This was followed by prepared discussion by: professors Elmer B. Mode, Boston University; Helen M. Walker, Teachers College, Columbia University; Samuel A. Stouffer, Harvard University, and Albert E. Waugh, Department of Economics, University of Connecticut. Many members participated in the general discussion. At the conclusion of this session, a film on Modern Quality Control was shown by Mr. Simon Collier of the Johns Manville Company.

Two Monday sessions, also held jointly with the American Statistical Association, and with the cooperation of the Operations Evaluation Group of the Navy and the Operations Analysis of the Air Force, were devoted to *Operations Research*. Professor Edward L. Bowles of Massachusetts Institute of Technology presided at the Morning session. The following papers:

1. *Operations Research in the Department of the Navy.*
   Dr. J Steinhardt, Director, Operations Evaluation Group.
2. *Operations Research in the Department of the Air Forces.*
   Dr. Leroy A. Brothers, Chief, Operations Analysis.

were followed by discussion by Dr. Arthur A. Brown, Operations Evaluation Group, Dr. Thomas I. Edwards, Operations Analysis, Professor G. Baley Price, The University of Kansas and Wartime Operations Analyst and Dr. W. J. Youden, Douglas Aircraft Company and Wartime Operations Analyst.

Dr. Merrill M. Flood, Assistant Deputy Director of Research and Development, General Staff, U. S. Army, presided at the afternoon session. The following papers were presented:

1. *Operations Analysis in the Southwest Pacific Air War.*
   Dr Roger I. Wilkinson, Bell Telephone Laboratories and Wartime Operations Analyst
2. *Operations Analysis of Air-Sea Rescue.*
   Dr. E S. Lamar, Operations Evaluation Group.
3. *Factorial Chi-Square in Test Shooting.*
   Dr. A. E. Brandt, Technical Director, Naval Ordnance Laboratory and Wartime Operations Analyst.
4. *Mathematical Techniques of Program Planning.*
   Dr George Dantzig, Consultant to the Air Comptroller, Headquarters, USAF.

A session on the *Application of the Theory of Extreme Values* was held jointly with the American Statistical Association on Tuesday, December 30. Professor Jacob Wolfowitz of Columbia University presided at the session. The following papers were presented:

1 *Introduction. The Mathematical Theory of Extreme Values.*
Professor Richard Von Mises, Harvard University.
2 *Applications to the Prediction of Flood Flows*
Professor Emil Gumbel, Brooklyn College.
3 *Applications to Meteorology*
Dr Horace Norton, Weather Bureau, Washington, D. C
4 *Applications to Fracture Problems*
Dr Benjamin Epstein, Coal Research Laboratory, Carnegie Institute of Technology.

The session concluded with discussion by Miss Marion Sandomire, Navy Department, Bureau of Ships and Dr Bradford Kimball, Port Washington, New York.

A session on *Statistical Techniques in Life Insurance* was held jointly with the American Statistical Association at 1:15 P M., December 30. Mr. Robert J. Myers, Actuarial Consultant, Social Security Administration, was chairman of the meeting The following papers were presented:

1. *Problems with Sampling Procedures for Reserve Valuations*
Mr George C Campbell, Supervisor, Actuarial Division, Metropolitan Life Insurance Company
2. *Sampling Errors in Life Insurance Mortality and Other Statistics.*
Mr Donald Cody, Assistant Actuary, Equitable Life Assurance Society
3. *Recent Developments in Graduation and Interpolation*
Dr. T. N. E Greville, National Office of Vital Statistics, U S Public Health Service.

A session of contributed papers was held at 3:30 P.M. on December 30. Dr. T. N. E. Greville of the National Office of Vital Statistics presided. The following papers were presented:

1. *Distribution of the Circular Serial Correlation Coefficient for Residuals from a Fitted Fourier Series (Preliminary Report)*
Professor R. L. Anderson, North Carolina State College and Professor T. W. Anderson, Jr., Columbia University.
2. *Some New Methods for Distributions of Quadratic Forms*
Professor Harold Hotelling, Institute of Statistics, University of North Carolina.
3. *Frequency Functions Defined by the Pearson Difference Equation.*
Professor Leo Katz, Michigan State College, East Lansing.
4. *Distribution of the Sum of Roots of a Determinantal Equation Under a Certain Condition*
Mr D. N. Nanda, Institute of Statistics, University of North Carolina.
5 *Applications of Carnap's Probability Theory to Statistical Inference.*
Professor Gerhard Tintner, Department of Economics, Iowa State College
6. *Circular Probable Error of an Elliptical Gaussian Distribution*
Dr. H. H. Germond, S. W. Marshall & Co., Washington, D. C.

The annual business meeting of the Institute was held at 4:30 P.M., December 29, 1947 in the ball room of the Commodore Hotel. There were reports by the President, Secretary-Treasurer, Mr Morris Hansen, Chairman of the Committee on Planning and Development, and Dr. John Curtiss, Chairman of the Program Committee Mr Hansen presented a tentative form of the proposed new constitution while Dr. Curtiss discussed program plans. There was some discussion on these general questions from the floor

Professor A. Wald was elected President, and Dr. Churchill Eisenhart and Professor Henry Scheffé, Vice-Presidents.

PAUL S. DWYER,
Secretary.

REPORT ON THE CHICAGO MEETING OF THE INSTITUTE

The thirty-second meeting of the Institute of Mathematical Statistics was held at the Sherman Hotel, Chicago, Monday and Tuesday, December 29–30. The meeting was held in conjunction with the one hundred fourteenth meeting of the American Association for the Advancement of Science and Co-operating Associated Societies. The following twenty-eight members of the Institute attended the meeting:

W. Bartky, D H. Blackwell, G. M Brown, I. W. Burr, A. G. Carlton, M. Castellanos, C W. Cotterman, A T Craig, J. H. Davidson, R. C. Davis, W. E Deming, M. Elveback, M L Garbuny, W W. Gutzman, T. J. Jaramillo, E. S. Keeping, T. C. Koopmans, E. L. Lahti, M M Lavin, K. May, J A. Pierce, O. Reiersol, H. Rubin, L. J. Savage, J. Silber, W. A Wallis, E. L Welker and J. W. Wilkins

The Monday afternoon session was devoted to contributed papers of Section A, AAAS, and of the Institute, and to the Vice-Presidential address of Section A. The following papers were presented:

1. *On the Boundary Layer Motion along a Periodically Oscillating Plane in Compressible Viscous Fluids.*
   Dr M Z. Krzywoblocki, University of Illinois.
2. *Variations of the Probability of Unfair Election Results.*
   Dr Kenneth May, Carleton College.
3. *Normal Equations with Nearly Vanishing Determinants.*
   Dr. M Herzberger and Dr. R. Norris
4. *Composition of Binary Quadratic Forms.*
   Professor Gordon Pall, Illinois Institute of Technology
5. *A Proof of the Asymptotic Analogue of the Theorem of Cramér and Rao.*
   Dr. Herman Rubin, Institute for Advanced Study.
6. *The Solution of Differential Equations in the Presence of Turning Points*, Vice-Presidential address of Section A.

The Tuesday afternoon session was also a joint session of Section A and the Institute, with Dean Walter Bartky of the University of Chicago presiding. The following two papers were presented upon invitation of the Institute:

1. *Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics.*
   Professor P R. Halmos and Dr. L J. Savage, University of Chicago.
2. *Unbiased Sequential Estimation.*
   ProfessorDavid Blackwell, Howard University.

REPORT OF THE PRESIDENT OF THE INSTITUTE FOR 1947

The healthy growth of the Institute has continued through 1947. The membership increased from 900 to 1046. This increase is gratifying as a sign that more and more people appreciate the usefulness of basic theory and are ready to support research by making our *Annals* possible. It is is also pleasing to note that statistical theory and methodology are reaching new fields and that new groups as a whole are becoming conscious of the usefulness of contact with mathematical statistics. These developments are reflected in the meetings of the Institute.

Meetings. The Ninth and Tenth Annual meetings (for 1946 and 1947) were held in the traditional way in conjunction with the meetings of the American Statistical Association (January—Atlantic City and Christmas—New York). The Tenth Summer Meeting was held with the American Mathematical Society and the Mathematical Association of America (September—Yale). Regional meetings were held in California (June—San Diego, December—Berkeley) and in Chicago (December), the latter in conjunction with the meetings of the American Association for the Advancement of Science (AAAS). Moreover, two meetings were organized with specialized programs of interest to groups with whom the Institute has not previously had much contact. A meeting in April at Columbia University, co-sponsored by the American Mathematical Society, was devoted to *Stochastic Processes and Random Noise*, and another meeting held simultaneously at Atlantic City was in conjunction with the meeting of the Eastern Psychological Association. It is clear that with such diversified meetings the Program Committee could not always act as a unit. J H. Curtiss was its Chairman and J. Neyman and J. W. Tukey arranged some of the programs. Other members of the Committee were: C. W. Churchman, T. Koopmans, F. C. Mosteller, J. Neyman, H. Scheffé, J. Wolfowitz, and H. Working.

At the Tenth Summer Meeting A. Wald delivered the first Henry L. Rietz Memorial Lecture. It is desirable to preserve the solemnity of the occasion of the Rietz lectures and it was therefore decided that they should not be given every year. Accordingly, no Rietz lecturer has been selected for 1948.

The Institute had no share in the program of the International Statistical Congress in Washington. However, Fellows of the Institute were invited to that Congress. This Congress and the Princeton Bi-Centennial were beneficial by establishing more intimate personal ties with our European colleagues. It is widely felt on both sides of the ocean that a closer cooperation, in particular with British statisticians, is highly desirable. Various suggestions in that direction were informally discussed in Washington and Princeton and M. G. Kendall has kindly consented to explore the practical possibilities. It is needless to say that the Institute is eager to do everything possible to promote cooperation and increase its usefulness also to our British colleagues.

Relations with other organizations It is gratifying to note that the cooperation of the Institute with sister societies is growing in intensity. The last two Presidential reports mentioned plans for a reorganization of the American Statistical Association with a view to more intimate relations among statistical societies. The revision of the constitution of the Association is not yet completed. It appears now that also the American Mathematical Society feels the need of closer collaboration with all groups interested in applied mathematics. It is too early to predict the results of these movements but it is clear that we must devote careful thought to our own organization and to our future relations with other groups.

In 1947 the AAAS organized an Inter-Society Committee for the National Science Foundation Legislation. At the first meeting in Washington we were represented by J H. Curtiss and W. A. Shewhart and at the meeting in December in Chicago by W. Bartky. In ballots on the two controversial subjects the Institute voted against exclusion of social sciences and abstained on the question of patent rights. W. Feller represented the Institute on the Policy Committee of the American Mathematical Society. Through this Committee the Institute went on record as favoring the National Science Foundation Bill. Otherwise the discussions of the Policy Committee were mostly connected with the establishment of an International Mathematical Union. Cletus O. Oakley represented the Institute on the Publicity Committee of the American Mathematical Society of which he is chairman. G. W. Snedecor was our representative on the AAAS Council, W. Bartky on the National Research Council, F. C. Mosteller and S S. Wilks on the Joint Committee for the Development of Statistical Application in Engineering and Manufacturing. In recent years the common interests of the Institute and the *actuarial profession* have grown in importance and it has been suggested that closer cooperation would be beneficial to both parts. A new committee has been established to explore these possibilities and in particular to arrange a joint meeting during 1948. Members of this committee are: G. C Campbell, T. N. E. Greville, C. Fisher, C. Spoerl, Chairman.

Internal Work. The growth of the Institute has rendered parts of the Constitution obsolete and a revision seems indicated In particular, it appears that the present system of elections is no longer satisfactory. The Institute is deeply indebted to its Committee on Planning and Development which has devoted much thought and consideration not only to a revision of the Constitution but also to the future development of the Institute as a whole. The membership had occasion to discuss the preliminary plans at two business meetings. M. H. Hansen acted as Chairman of the Committee; other members were: J. H. Curtiss, W. G Cochran, J. Neyman, H. W. Norton, F. F. Stephan, J. W. Tukey, W. A. Wallis.

A sharp increase in printing costs has, unfortunately, necessitated an increase in membership dues However, the membership should rest assured that the financial position of the Institute is intrinsically sound. The cash prospects for 1948 are not rosy, but this is due principally to the necessity of reprinting

back-numbers of the *Annals* which in itself is a sign of health and promise of stability At present the Institute has a considerable reserve in back numbers and this reserve is rapidly being transformed into cash. We are also exploring the possibilities of new revenue and have started a campaign to get advertisements for the *Annals*. A possible campaign for institutional members is held in abeyance pending a clarification of our formal relations with sister societies. In order to make the Annals available in European countries with monetary exchange restrictions, the dues and subscriptions have been increased only for the Western Hemisphere. The investments of the Institute have been supervised by the Finance Committee consisting of C. F. Roos, L. A. Knowler, F. F. Stephan, and Paul S Dwyer, Chairman

Last year's Committee on Teaching completed its work and submitted a detailed report which will be of great value It will be published in the *Annals of Mathematical Statistics*. The Committee has been dissolved with special thanks of the Board of Directors for their successful work. H. Hotelling was chairman and its members were Walter Bartky, W Edwards Deming, Milton Friedman, and Paul Hoel The Committee on Tabulation under the chairmanship of C. Eisenhart and consisting of Paul S Dwyer, H. Goldstine, A Lowan, H. W. Norton, and G R Stibitz has outlined the work for the coming years which promises to be of great interest

The Membership Committee consisted of C. C. Craig, P G. Hoel, and J. H. Curtiss as Chairman. On its recommendations the following members were elected Fellows: T. W. Anderson, David Blackwell, Frederick Mosteller, Gerhard Tintner, Charles P. Winsor, Alexander Aitken, George Darmois, Ragnar Frisch, Robert C. Geary, and John Wishart. The Nominating Committee consisted of Meyer A. Girshick, Paul G. Hoel, Horace W Norton, Frederick Mosteller, and George W. Snedecor, Chairman A. Wald was nominated for President, and as an innovation four nominations for Vice-presidents were made: C. Eisenhart, A. M. Mood, Henry Scheffé, F F. Stephan.

The *Annals of Mathematical Statistics* are covered by a special report of the Editor. However, it is appropriate to say that the Institute takes pride in the development of the *Annals*. While members see only its spectacular success, they should bear in mind that this is mostly due to the work of one man, S S. Wilks. In view of the great variety of interests of our membership and the many desirable directions in which the *Annals* could develop, it is clear that the work of the Editor can not always be pleasing and naturally often means a nervous burden. I feel sure that I speak for all our members in expressing the Institute's sincere thanks to S. S. Wilks not only for his work but also for his wisdom in striking a sensible balance between many wishes and possibilities and leading the *Annals* so successfully in a direction satisfactory to all of us.

In thanking all other members who have contributed to the work of the Institute, it is hard to find appropriate words to express appreciation for the unselfish efforts and devotion of our Secretary-Treasurer. Few members will realize how much of Dwyer's time and thoughts are spent for the Institute

and how much the smooth running of the affairs of the Institute is due to his hard work.

Finally, it is a pleasant duty to express our thanks and appreciation to Princeton University and to the University of Michigan. These Institutions have generously provided office space and other help which has greatly facilitated our work and saved us expenses.

WILL FELLER,
President, 1947.

December 31, 1947.

REPORT OF THE SECRETARY-TREASURER OF THE INSTITUTE FOR 1947

At the beginning of 1947 the Institute had 900 members and during 1947, 210 new members (10 of which begin their membership with 1948) joined the Institute. During 1947 the Institute lost 73 members, 43 by resignation, 25 by suspension for non-payment of dues, and 5 by death. The Institute has 1,037 members as it starts 1948.

The following members died during the year:

Margaret J Dix
Professor Irving Fisher
Albert M Freeman
Professor Henry A Ruger
Professor James G Smith

A summary of the financial transactions of the Institute is given in the *Financial Statement for 1947* which follows:

FINANCIAL STATEMENT

December 31, 1946 to December 31, 1947

A Receipts

Balance on Hand,* December 31, 1946		\$7,241 55
Dues		5,054.43
Life Membership Payments		287 50
Subscriptions		2,892 93
Sale of Back Numbers		3,969.95
Net Income from Investments		63 00
Miscellaneous		76.56
Total		\$19,585 92

B Expenditures

Annals—Current		
Office of Editor	\$160 40	
Waverly Press	7,145.79	\$7,306.19
Annals—Back Numbers		
Reprinted 500 copies each Vol III #1 & 2; IV #2; V #2, VII #4; XI #1 & 4; XII #1, XIV #1, 2 & 3	3,039 00	
Iowa City Office	143 75	3,182.75
Mathematical Reviews and Inter-Society for National Science Foundation		135.00

* In bank deposits and government bonds.

OFFICE OF THE SECRETARY-TREASURER		
Printing, memoranda, etc (including some stamped envelopes)	1,100.49	
Postage, supplies, express, telephone calls and cables	400.00	
Clerical help	1,502.31	3,002.80
MISCELLANEOUS		100.81
BALANCE ON HAND,* DECEMBER 31, 1947		5,858.37
TOTAL		$19,585.92

C. SUMMARY OF RECEIPTS AND EXPENDITURES

BALANCE ON HAND,* DECEMBER 31, 1946	$7,241.55
RECEIPTS DURING 1947	12,344.37
EXPENDITURES DURING 1947	13,727.55
BALANCE ON HAND,* DECEMBER 31, 1947	5,858 37

D. COMPARISON OF ASSETS ON DECEMBER 31, 1946 AND DECEMBER 31, 1947

	'46	'47
U S Government G Bonds	$5,000.00	$3,000.00
Life Membership Funds	1,888.00	1,888.00—Bonds
	139.50	427.00—Bank Dep.
Additional Bank Deposits	214.05	543.37
Current Accounts Receivable	452.62	423.55
Estimated Value (Cost) of back issues of *Annals***	7,234.58	10,866.73
TOTAL	$14,928.75	$17,148.65
Net Gain 1947		2,219.90

E LIABILITIES OF INSTITUTE OF MATHEMATICAL STATISTICS AS OF DECEMBER 31, 1947

All bills which have been presented have been paid. The Life Membership Fund now contains $2,315 00 which covers 30 members. Also $3,348.11 has been paid in for 1948 (and later) dues and subscriptions

The increase in the size of the *Annals* from 500 to 600 pages and the phenomenal activity in the sales of back numbers are the two most important factors to be considered in comparing the 1947 statement with those of previous years. The Waverly Press bills for 1946 totalled $4,566.27 while the corresponding amount for 1947 was $7,145.79 an increase of 56%. The increase is attributable not only to the increased size of the *Annals* but also to the fact that printing costs are rising rapidly and, to a less extent, to the fact that we are printing a larger number of copies. It is to be noted that the cost of the *Annals* alone in 1947 was over $2,000 more than the amount received from dues. As a result of the increase in dues, the 1948 report should be more satisfactory in this respect.

The phenomenal sales in back issues, noted in the report for 1946, were accelerated in 1947. We sold nearly $4,000 of back issues. These extensive sales were embarrassing to our cash position since they exhausted many of our issues and the continued reprinting forced us to place a considerable portion of

** Cost of *Annals* calculated at 67 cents per copy.

our reserves in inventory (some of which probably will not be returned to cash within decades). Eleven issues were reprinted during the first six months of 1947 The resulting low cash position forced a temporary change in the policy of reprinting issues as they became exhausted

It was necessary to cash two $1000 interest bearing G bonds to meet the Waverly and reprinting bills as they came due. These brought $1938.00 rather than $2000 as they have been valued in previous reports. As the income from bonds during the year was $125, I have entered the net income from investments as $63.00

An attempt has been made to keep down the costs of the office of the Secretary-Treasurer. The expense for 1947 was about $100 more than the expense for 1946 and seems very satisfactory in view of the larger membership and greatly increased costs of all materials and services.

For the reasons indicated above, the cash position (including bonds and Life Membership payments) was lowered during the year by $1,383.18. This is compensated for by an increase in the value of the stock of back issues (valued at cost) of $3,632.15. Some members of the finance committee feel that it is improper to list all of this stock as assets since we can probably sell only a portion of it in the next five or ten years. However, we did sell nearly $4,000 of *Annals* in 1947 and it is indicated (at the new prices) that the sales of issues we have now on hand will yield us $11,000 in the next five to ten years.

Many of the issues which were stored in Iowa City have been sold and Professor Knowler has sent the remaining issues to Ann Arbor. I wish to acknowledge the work of Professor Knowler in caring for these issues and to express the appreciation of the Institute for his efforts over a period of years. I also wish to express my appreciation to Mr. Carl Bennett who contributed much time and energy in looking after the back issues at Ann Arbor.

This report does not cover the amount of $390.20 which is held temporarily by the Institute for the fund for *Annals* for Countries Devastated by War. Arrangements are being made to purchase *Annals* for certain institutions which the Committee is recommending.

Paul S. Dwyer,
Secretary-Treasurer.

December 31, 1947.

REPORT OF THE EDITOR FOR 1947

During the past year the increase in the number of manuscripts submitted to the Annals has continued. More manuscripts have been received from foreign countries than in any preceding year. During 1947 papers were published by authors in Argentina, Australia, Canada, England, France and Sweden. If manuscripts continue to be received at the present rate it will not be possible to publish them in the *Annals* without further expansion. The gap between receipts of manuscripts and publication is likely to become serious by the end of 1948. The 1947 volume of the *Annals* contained 56 papers of which 25 were short notes. The total number of pages printed was 618, representing an increase of approximately 11% over the size of the 1946 volume. It now appears that increased printing costs will prevent a further increase in the size of the *Annals* for 1948. It is therefore extremely important that authors submitting papers to the *Annals* make every effort to keep their papers as brief as possible.

Contributions to probability and statistical theory are continuing to come in from a wide variety of fields. They were written by biologists, chemists, economists, mathematical statisticians, mathematicians and physicists, representing universities, government agencies and laboratories, business and industrial organizations. Some of these contributions are rather heterogeneous in quality of results and presentation. However, patient attempts are being made to have all papers with novel and interesting results suitably revised and published. Attempts to have expository papers prepared are being continued.

The Editor wishes to take this opportunity to acknowledge, on behalf of the Editorial Committee, the generous refereeing assistance which has been given by the following persons: L. A. Aroian, Z. W. Birnbaum, David Blackwell, A. H. Bowker, I. W. Burr, G. W. Brown, K. L. Chung, W. J. Dixon, T. N. E. Greville, F. E. Grubbs, J. B. S. Haldane, T. E. Harris, C. Hastings, L. Henkin, G. A. Hunt, B. F. Kimball, T. Koopmans, S. Kullback, E. L. Lehmann, H. Levene, H B. Mann, P. J. McCarthy, W. E. Milne, R. Otter, M. P. Peisakoff, H E. Robbins, L. J. Savage, F. F. Stephan, D. F. Votaw, and J. E. Walsh.

The Editor is also indebted to the following persons at Princeton University for preparation of manuscripts for the printer, and other editorial and office assistance: Miss Jacqueline G. Foster, M. F. Freeman and J. E. Walsh.

S. S. Wilks,
Editor.

December 31, 1947.

CONSTITUTION AND BY-LAWS OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Constitution

ARTICLE I

Name and Purpose

1 This organization shall be known as the Institute of Mathematical Statistics.

2. Its object shall be to promote the interests of mathematical statistics.

ARTICLE II

Membership

1. The membership of the Institute shall consist of Members, Fellows, Honorary Members, and Sustaining Members.

2. Voting members of the Institute shall be (a) the Fellows, and (b) all others, Junior members excepted, who have been members for twenty-three months prior to the date of voting.

3. No person shall be a Junior Member of the Institute for more than a limited term as determined by the Committee on Membership and approved by the Board of Directors.

ARTICLE III

Officers, Board of Directors, and Committee on Membership

1. The Officers of the Institute shall be a President, two Vice-Presidents, and a Secretary-Treasurer. The terms of office of the President and Vice-Presidents shall be one year and that of the Secretary-Treasurer three years. Elections shall be by majority ballots at Annual Meetings of the Institute. Voting may be in person or by mail.

(a) Exception. The first group of Officers shall be elected by a majority vote of the individuals present at the organization meeting, and shall serve until December 31, 1936.

2. The Board of Directors of the Institute shall consist of the Officers, the two previous Presidents, and the Editor of the Official Journal of the Institute.

3. The Institute shall have a Committee on Membership composed of a Chairman and three Fellows. At their first meeting subsequent to the adoption of this Constitution, the Board of Directors shall elect three members as Fellows to serve as the Committee on Membership, one member of the Committee for a term of one year, another for a term of two years, and another for a term of three years. Thereafter the Board of Directors shall elect from among the Fellows one member annually at their first meeting after their election for a term of three years. The president shall designate one of the Vice-Presidents as Chairman of this Committee.

ARTICLE IV

Meetings

1. A meeting for the presentation and discussion of papers, for the election of Officers, and for the transaction of other business of the Institute shall be held annually at such

time as the Board of Directors may designate. Additional meetings may be called from time to time by the Board of Directors and shall be called at any time by the President upon written request from ten Fellows. Notice of the time and place of meeting shall be given to the membership by the Secretary-Treasurer at least thirty days prior to the date set for the meeting. All meetings except executive sessions shall be open to the public. Only papers accepted by a Program Committee appointed by the President may be presented to the Institute

2. The Board of Directors shall hold a meeting immediately after their election and again immediately before the expiration of their term. Other meetings of the Board may be held from time to time at the call of the President or any two members of the Board. Notice of each meeting of the Board, other than the two regular meetings, together with a statement of the business to be brought before the meeting, must be given to the members of the board by the Secretary-Treasurer at least five days prior to the date set therefor Should other business be passed upon, any member of the Board shall have the right to reopen the question at the next meeting.

3 Meetings of the Committee on Membership may be held from time to time at the call of the Chairman or any member of the Committee provided notice of such call and the purpose of the meeting is given to the members of the Committee by the Secretary-Treasurer at least five days before the date set therefor. Should other business be passed upon, any member of the Committee shall have the right to reopen the question at the next meeting Committee business may also be transacted by correspondence if that seems preferable.

4 At a regularly convened meeting of the Board of Directors, four members shall constitute a quorum. At a regularly convened meeting of the Committee on Membership, two members shall constitute a quorum.

ARTICLE V

Publications

1. The *Annals of Mathematical Statistics* shall be the Official Journal for the Institute. The Editor of the *Annals of Mathematical Statistics* shall be a Fellow appointed by the Board of Directors of the Institute. The term of office of the Editor may be terminated at the discretion of the Board of Directors.

2. Other publications may be originated by the Board of Directors as occasion arises.

ARTICLE VI

Expulsion or Suspension

1. Except for non-payment of dues, no one shall be expelled or suspended except by action of the Board of Directors with not more than one negative vote.

ARTICLE VII

Amendments

1. This constitution may be amended by an affirmative two-thirds vote at any regularly convened meeting of the Institute provided notice of such proposed amendment shall have been sent to each voting member by the Secretary-Treasurer at least thirty days before the date of the meeting at which the proposal is to be acted upon. Voting may be in person or by mail.

By-laws

ARTICLE I

Duties of the Officers, the Editor, Board of Directors, and Committee on Membership

1 The President, or in his absence, one of the Vice-Presidents, or in the absence of the President and both Vice-Presidents, a Fellow selected by vote of the Fellows present shall preside at the meetings of the Institute and of the Board of Directors. At meetings of the Institute, the presiding officer shall vote only in the case of a tie, but at meetings of the Board of Directors he may vote in all cases. At least three months before the date of the annual meeting, the President shall appoint a Nominating Committee of three members. It shall be the duty of the Nominating Committee to make nominations for Officers to be elected at the annual meeting and the Secretary-Treasurer shall notify all voting members at least thirty days before the annual meeting. Additional nominations may be submitted in writing, if signed by at least ten Fellows of the Institute, up to the time of the meeting.

2 The Secretary-Treasurer shall keep a full and accurate record of the proceedings at the meetings of the Institute and of the Board of Directors, send out calls for said meetings and, with the approval of the President and the Board, carry on the correspondence of the Institute. Subject to the direction of the Board, he shall have charge of the archives and other tangible and intangible property of the Institute and upon the direction of the Board he shall publish in the *Annals of Mathematical Statistics* a classified list of all Members and Fellows of the Institute. He shall send out calls for annual dues and acknowledge receipt of same, pay all bills approved by the President for expenditures authorized by the Board or the Institute; keep a detailed account of all receipts and expenditures, prepare a financial statement at the end of each year and present an abstract of the same at the annual meeting of the Institute after it has been audited by a Member or Fellow of the Institute appointed by the President as Auditor. The Auditors shall report to the President.

3. Subject to the direction of the Board, the Editor shall be charged with the responsibility for all editorial matters concerning the editing of the *Annals of Mathematical Statistics*. He shall, with the advice and consent of the Board, appoint an Editorial Committee of not less than twelve members to co-operate with him; four for a period of five years, four for a period of three years, and the remaining members for a period of two years, appointments to be made annually as needed. All appointments to the Editorial Committee shall terminate with the appointment of a new Editor. The Editor shall serve as editorial adviser in the publication of all scientific monographs and pamphlets authorized by the Board

4. The Board of Directors shall have charge of the funds and of the affairs of the Institute, with the exception of those affairs specifically assigned to the President or to the Committee on Membership. The Board shall have authority to fill all vacancies ad interim, occurring among the Officers, Board of Directors, or in any of the Committees. The Board may appoint such other committees as may be required from time to time to carry on the affairs of the Institute The power of election to the different grades of Membership, except the grades of Member and Junior Member, shall reside in the Board.

5. The Committee on Membership shall prepare and make available through the Secretary-Treasurer an announcement indicating the qualifications requisite for the

different grades of membership. The Committee shall review these qualifications periodically and shall make such changes in these qualifications and make such recommendations with reference to the number of grades of membership as it deems advisable. The power to elect worthy applicants to the grades of Member and Junior Member shall reside in the Committee, which may delegate this power to the Secretary-Treasurer, subject to such reservations as the Committee considers appropriate. The Committee shall make recommendations to the Board of Directors with reference to placing members in other grades of membership. The Committee shall give its attention to the question of increasing the number of applicants for membership and shall advise the Secretary-Treasurer on plans for that purpose.

ARTICLE II

Dues

1. Members shall pay seven dollars at the time of admission to membership and shall receive the full current volume of the Official Journal. Thereafter, Members and Fellows shall pay seven dollars annual dues. Honorary members shall be exempt from all dues.

A Sustaining Member shall pay annual dues of a multiple of one hundred dollars.

An approved nominee of a Sustaining Member shall be a member in good standing without payment of dues for each year in which he is nominated provided that in that year he has been a member for less than three years

(a) Exception. In the case that two Members of the Institute are husband and wife and they elect to receive between them only one copy of the Official Journal, their dues shall each be reduced by twenty-five per cent.

(b) Exception. Any Member or Fellow may make a single payment which will be accepted by the Institute in place of all succeeding annual dues and which will not otherwise alter his status as a Member or Fellow and will be based upon a suitable table and rate of interest, to be specified by the Board of Directors.

(c) Exception. Any Member of Fellow of the Institute serving, except as a commissioned officer, in the Armed Forces of the United States, or of a friendly power, will, upon notification to the Secretary-Treasurer, be excused from the payment of dues until the January first following his discharge from service or his commissioning as an officer. He shall have all privileges of membership except that he shall not receive the Official Journal. However, during the first year of his resumed membership he may elect to receive one copy of each volume of the Official Journal published during the period of his service membership by paying one-half of the total of dues excused.

(d) Exception. Anyone who resides outside the Western Hemisphere shall pay five dollars annual dues.

2. Annual dues shall be payable on the first day of January of each year.

3. Five dollars of the annual dues of each Member and Fellow shall be for a subscription to the Official Journal Fifteen dollars of the dues of each Sustaining Member shall be for two subscriptions to the Official Journal, and the binding of one copy.

4 For each one hundred dollars of annual dues, a Sustaining Member shall be entitled to nominate two persons for membership in the Institute.

5. It shall be the duty of the Secretary-Treasurer to notify by mail anyone whose dues may be six months in arrears, and to accompany such a notice by a copy of this article. If such person fail to pay such dues within three months from the date of mailing such notice, the Secretary-Treasurer shall report the delinquent to the Board of Directors.

The Board of Directors may strike the delinquent's name from the rolls and withdraw all privileges of membership, and may reinstate the delinquent upon payment of arrears of dues.

ARTICLE III

Salaries

1. The Institute shall not pay a salary to any Officer, Director, or member of any committee

ARTICLE IV

Amendments

1. These By-Laws may be amended in the same manner as the Constitution or by a majority vote at any regularly convened meeting of the Institute, if the proposed amendment has been previously approved by the Board of Directors

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIX, No. 2 — June, 1948

TI

OF MATHE

DISCRIMINANT FUNCTIONS WITH COVARIANCE

By W. G. Cochran and C. I. Bliss

North Carolina State College; Connecticut Agricultural Experiment Station and Yale University

1. Summary. This paper discusses the extension of the discriminant function to the case where certain variates (called the covariance variates) are known to have the same means in all populations. Although such variates have no discriminating power by themselves, they may still be utilized in the discriminant function.

The first step is to adjust the discriminators by means of their 'within-sample' regressions on the covariance variates. The discriminant function is then calculated in the usual way from these adjusted variates. The standard tests of significance for the discriminant function (e.g. Hotelling's T^2 test) can be extended to this case without difficulty. A measure is suggested of the gain in information due to covariance and the computations are illustrated by a numerical example. The discussion is confined to the case where only a single function of the population means is being investigated.

2. Introduction. Discriminant function analysis is now fairly well advanced for the case where there are only two populations. The data consist of a number of measurements, called the *discriminators*, that have been made on each member of a random sample from each population. The technique has various uses. Fisher [1] used it in seeking a linear function of the measurements that could be employed to classify new observations into one or other of the two populations. He pointed out [2] that a test of significance of the difference between the two samples, developed from his discriminant, was identical with Hotelling's generalization of Student's t test, discovered some years earlier [3]. Mahalanobis' concept of the generalized distance between two populations [4] was also found to be closely related to the discriminant function. In any of these applications—to classification, testing significance, or estimating distance—we may also be interested in considering whether certain of the measurements really contribute anything to the purpose at hand, and helpful tests of significance are available for this purpose.

Recently the authors encountered a problem in which it seemed advisable to combine discriminant function analysis with the analysis of covariance. This case occurs whenever, in addition to the discriminators, there is a measurement whose mean is known to be the same in both populations. Suppose, for example, that the I.Q.'s of each of a sample of students are measured. The sample is then divided *at random* into two groups, each of which subsequently receives a different type of training. Measurements made at the end of the period of training would be potential discriminators, but in the case of the initial I.Q.'s we can

clearly assume that there is no difference in the means of the populations corresponding to the two groups.

The initial I.Q. measurements are of course of no use in themselves in studying differences introduced by the training. Nevertheless, if they are correlated with the discriminators, they may serve in some way to 'improve' the discriminant: e.g. to increase the power of Hotelling's T^2 test, or to reduce the number of errors in classification. This paper discusses the problem of utilizing such measurements, which will be called *covariance variates.* The problem is analogous to that which is solved by the analysis of covariance. In covariance, as applied for instance in a controlled experiment, variates that are unaffected by the experimental treatments can be used to provide more accurate estimates of the effects of the treatments or to increase the power of the F test of the differences among the treatment means.

The procedure suggested is as follows. First, the multiple regression is obtained of each discriminator on all the covariance variates. These regressions are calculated from the 'within-sample' sums of squares and products: that is, from the sums of squares and products of deviations of the individual measurements from their sample means. Each discriminator is then replaced by its deviations from the multiple regression, and a new discriminant function is calculated in the usual way from these deviations. The extensions of Hotelling's T^2 and Mahalanobis' distance are both obtained from this discriminant, though a further adjustment factor is needed for tests of significance.

This paper is arranged in three parts. Part I presents a numerical example. The decision to place the example first was taken because most of the actual applications of the discriminant function in the literature appear to have been made by persons relatively unfamiliar with the theory of multivariate analysis. It is hoped that with the aid of the example readers in this class may be able to utilize covariance variates. For the same reason, the calculations have been presented as far as possible in terms of the operations of ordinary multiple regression, rather than in the form in which they first emerge from the theory. Actually, various equivalent methods of calculation are available, and it is not claimed that our method is necessarily the best. A mathematical statistician may prefer to follow the computing methods which come directly from theory (Part II, section 13).

The example is more complex in structure than the two-sample case. The data constitute a two-way classification, in which the row means are nuisance parameters, being of no interest, while only a single linear function of the column means is of interest. It is well known that the ordinary t test can be applied not only to the difference between two sample means, but to any linear function of a number of sample means in data that are quite complex. Discriminant function technique can be extended in the same way, and readers familiar with the analysis of variance should find no great difficulty in making the appropriate extension to such data.

Part II presents the theory. The reader who is primarily interested in theory

should read Part II before Part I. Since the approaches used by Mahalanobis, Hotelling and Fisher all converge, we have chosen that of Mahalanobis, mainly because the extension of his techniques to include covariance variates seems straightforward. Maximum likelihood estimation of the generalized distance is presented in full for the two-population case. The frequency distribution of the estimated distance and the extension of the T^2 test are worked out. An attempt is also made to obtain a quantity that will measure what has been gained by the use of covariance.

In order to illustrate how the theory applies with other types of data, the mathematical model is given for the row by column classification that occurs in the example. The major results for this model are indicated, though without proof.

In Part III it is shown that the computational methods used in the example are equivalent to those developed by theory. While this can easily be verified in a particular case, it is not intuitively obvious.

PART I NUMERICAL EXAMPLE

3. Description. The data form part of an experiment on the assay of insulin of which other parts have been published [5]. Twelve rabbits were used. Each rabbit received in succession four doses of insulin, equally spaced on a log. scale. An interval of eight days or more elapsed between successive doses, and the order in which the doses were given to any rabbit was determined by randomization. Thus the experiment is of the 'randomized blocks' type, where each rabbit constitutes a block and there are 12 blocks with 4 treatments each.

The effect of insulin is usually measured by some function of the blood sugar of the rabbit in periodic bleedings after injection of the insulin. The blood sugar was measured for each rabbit at 1, 2, 3, 4, and 5 hours after injection, and also before injection. In order to simplify the arithmetic, only the initial blood sugar and the blood sugars at 3 and 4 hours after injection will be considered here. These data are shown for the first three rabbits (with totals for all 12 rabbits) in Table I.

Let x_{iwz} be a typical observation of blood sugar, where $i = 3, 4$ stands for the hour after injection, w for the rabbit and d for the dose. The mathematical model to be used is as follows.

$$x_{iwz} = \mu_i + \rho_{iw} + \gamma_{iz} + \beta_{i0}(x_{0wz} - x_{0\cdot\cdot}) + e_{iwz}\,. \tag{1}$$

The parameters μ_i, ρ_{iw} and γ_{iz} represent the true mean and the effects of rabbit and log dose respectively. The quantity x_{0wz} is the initial blood sugar for the rabbit w before the test at dose z, while $x_{0\cdot\cdot}$ is the average initial blood sugar over the whole experiment. The blood sugar at i hours has been found experimentally to be correlated with the corresponding initial blood sugar, and the relationship is represented here as a linear regression, with β_{i0} as the regression coefficient. The residuals e_{iwz} are assumed to follow a multivariate (in this case bivariate) normal distribution, with zero means. The covariance between e_{iwz} and e_{jwz}

is taken as $\sigma_{ij\cdot 0}$. The model is the standard one for the ordinary analysis of covariance, except that we have *two* measures of the effect of insulin, x_3 and x_4.

One additional assumption was made. For all post-injection readings, the blood sugar seemed linearly related to the log dose l_s. Since this result has been found in other experiments, we assumed that

$$\gamma_{is} = \delta_i l_s$$

where δ_i is the regression coefficient of blood sugar on log dose.

4. Object of the analysis. Our object was to find the linear combination of the three blood sugar readings that would measure best the effect of the insulin. Because of the linearity of the regression on log dose, the effect of insulin on each

TABLE 1

Sample of original data on blood sugar levels in insulin experiment

Rabbit No.	Log dose											
	Initial blood sugar x_0				Three hours x_3				Four hours x_4			
	.32	.47	.62	.77	.32	.47	.62	.77	.32	.47	.62	.77
1	75	94	107	94	95	76	67	56	96	95	115	91
2	91	86	83	93	98	90	77	69	104	87	90	89
3	97	99	90	91	84	76	59	48	93	102	85	90
Total*.	1065	1074	1121	1070	932	872	731	591	1098	1020	970	847

*12 rabbits.

x_i is known completely if the slope δ_i is known. It seems reasonable to choose the linear compound of the x_i's which will give the maximum ratio when its estimated regression on log dose is divided by the estimated standard error of this regression We now consider how to obtain this maximum. The argument given below is not intended to prove the validity of the method, for which reference should be made to Part II.

The true regression of the original blood sugar x_0 on log dose is known to be zero. Hence, it is clear that the variate x_0 is useful only in so far as it enables us to obtain more accurate estimates of δ_3 and δ_4. For this purpose we need to estimate the effect of x_0 upon x_3 and x_4, the blood sugar readings at 3 and 4 hours, independently of dose of insulin or of differences between rabbits. From the standard theory of covariance the best estimate is the regression coefficient $b_{i0} = E_{i0}/E_{00}$, where E denotes a sum of squares or products calculated from the *error* line in the analysis of covariance, that is from the sums of squares and products of deviations of the x_i from the fitted regression on row and column parameters.

The regression of the blood sugar at each hour on the log dose of insulin is calculated from totals adjusted for the regression on x_0. Since the 4 successive log doses ($z = 1, 2, 3, 4$) are spaced equally, they may be replaced in the computation by the coded doses -3, -1, $+1$, and $+3$. If we let T_{iz} be the total blood sugar, summed over 12 rabbits, at the ith hour with dose z, the following result is well known for the analysis of covariance. The best estimate of $\delta_i (i = 3, 4)$ is

$$[(-3T_{i1} - T_{i2} + T_{i3} + 3T_{i4}) - b_{i0}(-3T_{01} - T_{02} + T_{03} + 3T_{04})]/240$$

The divisor, 240, is $12(3^2 + 1^2 + 1^2 + 3^2)$. The expression may be written

$$\frac{d'_i}{240} = \frac{(d_i - b_{i0} d_0)}{240},$$

where

$$d_i = -3T_{i1} - T_{i2} + T_{i3} + 3T_{i4}.$$

A linear combination is formed from d'_3 and d'_4, the numerators in the best estimates of δ_3 and δ_4, by means of the coefficients L_3 and L_4. L_3 and L_4 are computed so as to maximize the ratio of

$$d_I = L_3 d'_3 + L_4 d'_4$$

to its estimated standard error.

From the definition of d'_i, this requires a discriminant of the form

$$I = L_3(x_{3wz} - b_{30}x_{0wz}) + L_4(x_{4wz} - b_{40}x_{0wz}),$$

where each x_{0wz} is measured from its mean.

We require next the estimated standard error of d_I. This depends, in turn upon the variances of d'_3 and d'_4 and their covariance. As usual in the analysis, of variance we have

$$(5) \qquad V(d'_3) = V(d_3) + d_0^2 V(b_{30}) = \sigma_{33\cdot 0}\left(240 + \frac{d_0^2}{E_{00}}\right).$$

The residual variance $\sigma_{33.0}$ is estimated from the sums of squares and products in the error row of the analysis of covariance as

$$s_{33.0} = E_{33.0}/n = (E_{33} - E_{30}^2/E_{00})/n,$$

where n is the degrees of freedom in each E_{ii} diminished by one. Similar methods lead to the variance of d'_4 and to the covariance of d'_3 and d'_4. It follows that the *true* variance of d_I may be written

$$(6) \qquad V(d_I) \propto L_3^2\, \sigma_{33.0} + 2L_3L_4\, \sigma_{34.0} + L_4^2\, \sigma_{44.0},$$

where the factor $\left(240 + \frac{d_0^2}{E_{00}}\right)$ in equation (5) is omitted since it does not involve

the L's. Similarly, the *estimated* variance of d_I, apart from constant factors, may be written as

$$L_3^2E_{33.0} + 2L_3L_4E_{34.0} + L_4^2E_{44.0}\,. \tag{7}$$

The quantity to be maximized is therefore

$$\frac{(L_3 d_3' + L_4 d_4')}{\sqrt{L_3^2E_{33.0} + 2L_3L_4E_{34.0} + L_4^2E_{44.0}}}\,.$$

Formally, this is the same type of quantity that is maximized in ordinary analysis with the discriminant function. Differentiation with respect to the L's leads to the equations (after omission of another constant factor)

$$E_{33.0}L_3 + E_{34.0}L_4 = d_3'\,, \qquad E_{34.0}L_3 + E_{44.0}L_4 = d_4'\,. \tag{8}$$

The objective of the computation, therefore, is to obtain discriminant coefficients having the same ratio to each other as L_3 and L_4 in equations (8). As will be shown in the next section, this can be accomplished in practice more conveniently by substituting an alternative set of three simultaneous equations for the two in equations (8).

5. Calculations. The first step is to form the sums of squares and products in the analysis of covariance. With 12 rabbits and 4 doses, the conventional breakdown of each total sum is into components for rabbits (11 d.f.), doses (3 d.f.) and rabbits $\times$ doses (33 d.f.). Because of the assumed linear regression on log dose, the sum of squares for doses was further divided into two components The first (1 d.f.) is the contribution due to this regression. For x_i, the sum of squares due to regression is $d_i^2/240$, or in the case of x_3, $(1164)^2/240$, or 5645. The remaining component, (2 d.f.) is called the *curvature*, since it measures the effect of deviations from the linear regression. The sum of squares for curvature is found by subtraction.

The following points may be noted. (i) For both x_3 and x_4, the F ratio of the curvature mean square to the rabbits $\times$ doses mean square will be found to be less than 1, so that the data do not suggest rejection of the hypothesis of a linear regression on log dose. (ii) The F ratios of the regression mean squares to the rabbits $\times$ doses mean squares are highly significant, being 57.8 for x_3 and 28.7 for x_4. This indicates, incidentally, that the three-hour reading may be a more responsive measure of the effect of insulin than the four-hour reading. (iii) With x_0, the F ratio does not approach significance for either the regression or the curvature, as is to be expected.

A consequence of the assumption of linear regression on log dose is that the curvature mean squares and products are estimates of the same quantities as the rabbits $\times$ doses mean squares and products. Consequently, the lines for curvature and rabbits $\times$ doses in Table 2 could be added to give 35 d.f. for the 'error' sums of squares or products, E_{33}, etc. We decided, however, to estimate

the error only from the 33 d f. for rabbits × doses. This was done because it seemed to facilitate a test of the curvature of the final discriminant I. (This test will not be reported here.)

The L's could now be obtained from equations (8). In this case the first equation would contain the terms

$E_{33.0} = 3223 - (1259)^2/2351$; $E_{34.0} = 1200 - (1259)(1340)/2351$;

$$d_3' = d_3 - b_{30} d_0 = -1164 - \left(\frac{1259}{2351}\right) 62,$$

leading to the simultaneous equations

$$2548.8\,L_3 + 482.4\,L_4 = -1197.2$$

$$482.4\,L_3 + 2373.2\,L_4 = -844.3,$$

which give $L_3/L_4 = -.41848/-.27070 = 1.5459$.

TABLE 2

Sums of squares and products

Component	d f	x_0^2	x_3^2	x_4^2	x_0x_3	x_0x_4	x_3x_4
Between rabbits	11	886	9376	11165	1952	2477	9206
Between doses	3	168	5806	2810	−247	−98	3981
{Reg. on log dose	1	16	5645	2727	−301	−209	3924
{Curvature	2	152	161	83	54	111	57
Rabbits × doses	33	2351	3223	3137	1259	1340	1200
Total	47	3405	18405	17112	2964	3719	14387

Instead of using these equations, we propose to solve alternatively the set of three equations

$$\begin{aligned} S_{00}L_0 + S_{03}L_3 + S_{04}L_4 &= d_0 \\ S_{30}L_0 + S_{33}L_3 + S_{34}L_4 &= d_3 \qquad (9) \\ S_{40}L_0 + S_{43}L_3 + S_{44}L_4 &= d_4, \end{aligned}$$

where each S_{ij} ($i = 0, 3, 4$) is the sum of squares or products formed by adding the *error* line in the analysis of variance to the line for *regression on log dose.* Thus S_{ij} has 34 d.f. The ratio of L_3 to L_4, as found from equations (9), is exactly the same as that found from the original equations (8), as is proved in section 18. Further, the solution of the new equations seems to be more useful for performing tests of significance, as will appear in following sections

Accordingly, the first step after forming the analysis of variance is to set up the three equations (9).

The equations are solved by means of the inverse matrix. The values of d_i on the right side of the equations are replaced successively by 1, 0, 0 by 0, 1, 0 and by 0, 0, 1 to obtain the three sets of values for L_0, L_3 and L_4. These results are given in the first three columns of Table 4 and are designated as c_{ij}.

The L's follow from the c_{ij} by the usual rule for regressions. For example,

$$L_3 = \{(.003209)(62) + (.227781)(-1164) + (-.199655)(-809)\}\cdot 10^{-3} = -.103417$$

TABLE 3

Equations for determining L_3 and L_4

$$\begin{aligned} 2367L_0 + 958L_3 + 1131L_4 &= 62 \\ 958L_0 + 8868L_3 + 5124L_4 &= -1164 \\ 1131L_0 + 5124L_3 + 5864L_4 &= -809 \end{aligned}$$

The composite response or discriminant, adjusted for the covariance variate, is now taken as

$$I = L_3\left(x_3 - \frac{E_{30}}{E_{00}}\,x_0\right) + L_4\left(x_4 - \frac{E_{40}}{E_{00}}\,x_0\right)$$

or

$$-.103417\left(x_3 - \frac{1259}{2351}\,x_0\right) - .066883\left(x_4 - \frac{1340}{2351}\,x_0\right)$$

$$= .093503x_0 - .103417x_3 - .066883x_4\,.$$

Note that the value of L_0 is not used at this stage and that $L_3/L_4 = 1.546$ agrees with the value found from equations (8).

TABLE 4

Inverse matrix ($\times$ *10*^{3}) *and* L's

$(10^3 c_{ij})$			d_i	L_i
.465408	.003209	−.092568	62	.100008
.003209	.227781	−.199655	−1164	−.103417
−.092568	−.199655	.362846	−809	−.066883

A similar method may be followed when there are more discriminators or more covariance variates. With two covariance variates, x_0 and x_0^1, for instance, the adjusted discriminant would be

$$L_3(x_3 - b_{30}x_0 - b_{30}^1x_0^1) + L_4(x_4 - b_{40}x_0 - b_{40}^1x_0^1)$$

where b_{30}, b_{30}^1 are the partial regression coefficients of x_3 on x_0, x_0^1 respectively, determined from the error line, and similarly for x_4. Further, since any linear

function of the column (dose) means may be represented as a regression on some variate t_z, this method may be applied to any linear function of the column means in which we are interested, provided that the mathematical model is appropriate.

6. **Test of the regression of the adjusted discriminant on log dose.** The numerator of the regression of I on the coded doses is

$$d_I = L_3(d_3 - b_{30}d_0) + L_4(d_4 - b_{40}d_0).$$

Since the regressions of x_3 and x_4 on the coded doses were both significant, it may be confidently expected that the regression of I will also be significant. The test of significance will, however, be given in case it may be useful in other applications. For those who are familiar with multiple regression, the test is perhaps most easily made by means of a device due to Fisher [2].

Construct a dummy variate y_{wz} such that y_{wz} is always equal to t_z, or in our case to the coded doses. That is, y takes the value -3 for all observations at

TABLE 5

Analysis of y^2 and yx_i

	d f.	y^2	yx_i
Rabbits	11	0	0
Doses	3	240	d_i
Regression on log dose	1	240	d_i
Curvature	2	0	0
Rabbits $\times$ doses = error	33	0	0
Sum = Error plus reg. on log dose	34	240	d_i
Total	47	240	d_i

the lowest dose level, and -1, $+1$, and $+3$ respectively for all observations at the successive higher dosage levels. We shall show that equations (9) solved in finding the L's are formally the same as a set of normal equations for the linear regression of y on x_0, x_3, and x_4.

The following analysis for y^2 and yx_i may easily be verified.

It will be noted that the sum of products of y and x_i in the sum line is d_i. Further, S_{ij} is the sum of products of x_i and x_j for this line. It follows that the normal equations for the regression of y on the x's, as calculated from the "sum" line, are

$$S_{i0}L_0 + S_{i3}L_3 + S_{i4}L_4 = d_i \qquad (i = 0, 3, 4).$$

These are just the equations solved in obtaining the L's. Consequently, L_3 and L_4 are the partial regression coefficients of y on x_3 and x_4. A test of the null

hypothesis that the true values of L_3 and L_4 are both zero can be made by the standard method for multiple regression, as will be shown later from theory. This test is equivalent to a test of the hypothesis that the true value of d_l is zero.

To apply the test, we require three items in the analysis of variance of y. First, the total sum of squares for the Sum line, already seen to be 240 (Table 5). Second, the reduction due to a regression on all variates (covariance variates plus discriminators). By the usual rules for regression, this is (from Table 4)

$$L_0 d_0 + L_3 d_3 + L_4 d_4 = (.100008)(62) + (-.103417)(-1164) + (-.066883)(-809) = 180.69.$$

Finally, we need the reduction due to a regression on the variates that are not being tested, i.e. on the covariance variates alone. From Table 4, the reduction

TABLE 6

Analysis of variance of dummy variate y

	d. f.	S. S.	M. S.
Reduction to regression on covariance variates	1	1.62	
Additional reduction to regression on discriminators	2	179.07	89.54
Deviations	31	59.31	1.913
Total (from Sum line)	34	240.00	

in this case is simply d_0^2/S_{00} or $(62)^2/2367$, or 1.62. The difference, 180.69 − 1.62, represents the reduction due to the regression of y on L_3 and L_4, after fitting x_0. The resulting analysis is given below, the degrees of freedom being apportioned by the usual rules.

The F ratio, 89.54/1 913, or 46.80, with 2 and 31 d.f., is used to test the null hypothesis that the adjusted discriminant has no real regression on log dose.

7. Test of particular discriminators. Another useful test is that of the null hypothesis that a particular discriminator, or group of discriminators, contribute nothing to the adjusted discriminant. In other words, this is a test of the null hypothesis that the true values of a subset of the L's are all zero. The test is of interest in the present investigation, since it would be useful to know whether all five hourly readings of the blood sugar are really helpful. As might be expected by analogy with the previous section, the test is made by calculating the additional reduction due to the regression of y on the particular subset of the L's in question.

The test will be illustrated with respect to L_4. One method of making the test is to re-solve the normal equations with L_4 omitted. From this solution

the reduction due to a regression of y on x_0 and x_3 alone is obtained. The additional reduction due to a regression on x_4 is found by subtraction from 180 69.

However, the additional reduction can be found directly from the well-known regression theorem that it is equal to L_4^2/c_{44}. The c's have already been found in Table 4. The result is $(.066883)^2/(.000362846)$, or 12.33 This value is tested against the residual error mean square of 1.913, F having 1 and 31 d.f. The contribution is found to be significant.

In fact, by this process a kind of estimated standard error can be attached to each of the L's for the discriminators, using the formula $s\sqrt{c_{ii}}$, where s is the residual root mean square. Thus for L_3, (−.103417), the 'standard error' is $\sqrt{(1.913)(.000227781)}$, or .0209. It should be stressed that at this point the analogy with regression is rather thin. The L's are not normally distributed, nor do the estimated standard errors follow their usual distribution. It is, however, correct that if the true value of L_4 is zero, $L_4/s\sqrt{c_{44}}$ follows the t distribution with 31 d.f. Thus, if omission of some discriminators seems warranted,

TABLE 7

Analysis of variance for regression of y on the discriminators

	d f.	S S.	M. S
Regression.	2	159 20	79 60
Deviations.	32	80.80	2.525
Total.	34	240.00	

these t ratios are relevant in deciding which variate to eliminate first. Strictly speaking, the c's should be re-calculated after each elimination before deciding which other discriminators might also be discarded.

8. Estimation of the gain due to covariance. The tests given above enable us to state whether the discriminators contribute significantly, in the statistical sense. It is also of interest to investigate what has been gained by the use of the covariance variates. From the practical point of view, the question: "What is the gain from covariance?" might be re-phrased as: "If x_0 is ignored, how many rabbits must be tested in order to estimate the regression on log dose as accurately as it was estimated with the adjusted discriminant for 12 rabbits?"

The theoretical aspects of the question are discussed in section 16; the calculations are described here. The only new quantity needed is the F ratio for the regression of y on the discriminators alone. This can be obtained by a new solution of the normal equations, this time with the covariance variates omitted. With just one covariance variate, it is quicker to use the fact that the additional reduction to the regression of y on x_0, after fitting x_3 and x_4, is L_0^2/c_{00}, or $(.100008)^2/(.000465408)$ or 21.49. Consequently, the reduction due to a

regression of y on x_3 and x_4 alone is 180.69 − 21.49, or 159.20. The F ratio, 79 60/2.525, is 31.52, whereas the F ratio with covariance is 46.80 (from Table 6). The quantity suggested from theory for comparing the two techniques is

$$\frac{(n_2 - 2)F}{n_2} - 1$$

where n_2 is the number of d.f. in the denominator of F. These values are $\{(30 \times 31.52/32) - 1\}$ or 28.55 with no covariance and $\{(29 \times 46.80/31) - 1\}$ or 42.78 with covariance. The relative information is estimated as 42.78/28.55, or 1.50, so that the use of covariance gives 50 per cent more information. In other words about 18 rabbits would be needed if the initial blood sugars were ignored. To a slight extent this estimate favors the covariance analysis, since it ignores the increased accuracy that would accrue from the extra error d.f. if 18 rabbits were used without covariance.

PART II THEORY

9. Notation. The theory will be given first for the two-population case. We suppose that a random sample of size N has been drawn from each population. A typical discriminator is written $x_{iw\alpha}$ and a typical covariance variate $x_{\xi w\alpha}$, where

$i, j = 1, 2, \cdots p$ denote discriminators,

$\xi, \eta = 1, 2, \cdots k$ denote covariance variates,

$w = 1, 2$ denotes the population, and

$\alpha = 1, 2, \cdots N$ denotes the order within the sample.

The population mean of $x_{iw\alpha}$ is μ_{iw}, and the corresponding sample mean is $x_{iw\cdot}$. The difference $(\mu_{i2} - \mu_{i1})$ is denoted by δ_i and the corresponding estimated difference $(x_{i2\cdot} - x_{i1\cdot})$ by d_i.

10. Discriminant functions and generalized distance. Since we propose to approach the theory by means of the generalized distance, it may be well to review briefly the relation between the discriminant and the generalized distance. In the ordinary theory (with no covariance variates) it is assumed that the variates $x_{iw\alpha}$ follow a multivariate normal distribution, and that the covariance matrix σ_{ij} between $x_{iw\alpha}$ and $x_{jw\alpha}$ is the same in both populations. The generalized distance of Mahalanobis is defined by

$$p\Delta^2 = \sum_{i,j=1}^{p} \sigma^{ij} \delta_i \delta_j, \qquad \text{where} \qquad (\sigma^{ij}) = (\sigma_{ij})^{-1}. \tag{10}$$

In order to estimate this quantity from the sample, we first calculate the mean within-sample covariance s_{ij}, where

$$s_{ij} = \sum_{w=1}^{2} \sum_{\alpha=1}^{N} (x_{iw\alpha} - x_{iw\cdot})(x_{jw\alpha} - x_{jw\cdot})/2(N - 1), \tag{11}$$

The estimated distance is then taken as

$$pD^2 = \sum_{i,j=1}^{p} s^{ij} d_i d_j . \tag{12}$$

Apart from a factor $N/(N-1)$, this is the maximum likelihood estimate.

In the discriminant function used by Fisher (1), the object is to find a linear function $I_{w\alpha} = \Sigma M_i x_{iw\alpha}$, where the M_i are chosen to maximize the ratio of the sum of squares between samples to that within samples in the analysis of variance of I. This is equivalent to maximizing the ratio of the difference between the two sample means of I to the estimated standard error of this difference. As Fisher showed (2), the M_i (apart from an arbitrary multiplier) are given by

$$M_i = \sum_{j=1}^{p} s^{ij} d_j$$

Consequently, the difference between the two sample means of I, the discriminant function, is

$$\sum_{i=1}^{p} M_i d_i = \sum_{i,j=1}^{p} s^{ij} d_i d_j .$$

This is exactly the same as pD^2 in equation (12). Thus the discriminant function leads to the estimated distance, and *vice versa.*

11. Extension to the present problem. In our case there are $(p + k)$ variates (p discriminators, k covariance variates) from which to estimate the distance. All variates, $x_{iw\alpha}$ and $x_{\xi w\alpha}$, are assumed to follow a multivariate normal distribution. The covariance matrix, assumed the same in both populations, now has $(p + k)$ rows and columns, and may be denoted by

$$\Lambda = \begin{pmatrix} \sigma_{ij} & \sigma_{i\eta} \\ \sigma_{\xi j} & \sigma_{\xi\eta} \end{pmatrix}. \tag{13}$$

For each of the covariance variates, it is known that the population means $\mu_{\xi 1}$, $\mu_{\xi 2}$ are equal, so that the difference δ_ξ is zero. This is the fact that distinguishes the problem from ordinary discriminant function analysis.

Hence, the generalized distance, as defined from all $(p + k)$ variates contains no contribution from terms in δ_ξ and is given by

$$(p + k)\Delta^2 = \sum_{i,j=1}^{p} \sigma^{ij}_{(p+k)} \delta_i \delta_j . \tag{14}$$

The matrix $\sigma^{ij}_{(p+k)}$ is that formed by the first p rows and columns of the inverse of Λ. Note that in general this will not be the same as the matrix σ^{ij}, which is the inverse of σ_{ij}.

In the next section we consider the estimation of this quantity from the sample data. By analogy with the previous section, it might be guessed that the estimate would be of the form $\Sigma s^{ij}_{(p+k)} d_i d_j$. The maximum likelihood estimate

turns out to be of this form, except that instead of d_i we have d'_i, the difference between the two sample means of the deviations of x_i from its 'within-sample' linear regression on the x_ξ.

12. **Estimation of the distance.** It is known that the generalized distance is invariant under non-singular linear transformations of the variates. For convenience, we replace the $x_{iw\alpha}$ by variates $x'_{iw\alpha}$, where

$$x'_{iw\alpha} = x_{iw\alpha} - \sum_{\xi=1}^{k} \beta_{i\xi}(x_{\xi w\alpha} - \mu_{\xi w}).$$

Thus $x'_{iw\alpha}$ is the deviation of $x_{iw\alpha}$ from its population linear regression on the $x_{\xi w\alpha}$. The population mean of $x'_{iw\alpha}$ is clearly μ_{iw}, and the difference between the two population means is therefore δ_i.

The covariance matrix of the $x'_{iw\alpha}$, $x_{\xi w\alpha}$ may be written

$$\Lambda' = \begin{pmatrix} \sigma_{ij\cdot\xi} & 0 \\ 0 & \sigma_{\xi\eta} \end{pmatrix}, \tag{15}$$

where $\sigma_{ij\cdot\xi}$ denotes the covariance matrix of the deviations of the $x_{iw\alpha}$ from their regressions on the $x_{\xi w\alpha}$. It follows that in terms of the transformed variates the generalized distance is given by

$$(p + k)\Delta^2 = \sum_{i,j=1}^{p} \sigma^{ij\cdot\xi}\delta_i\delta_j \tag{16}$$

where $\sigma^{ij\cdot\xi}$ is the inverse of the $p \times p$ matrix $\sigma_{ij\cdot\xi}$.

The joint distribution of the $2N$ observations on each of the $x'_{iw\alpha}$ and $x_{\xi w\alpha}$ is as follows:

$$(2\pi)^{-N(p+k)} \mid \sigma^{ij\cdot\xi} \mid^{+N} \mid \sigma^{\xi\eta} \mid^{+N} \Pi dx'_{iw\alpha}\, dx_{\xi w\alpha}\cdot$$
$$\exp\left\{-\frac{1}{2}\left[\sum_{w=1}^{2}\sum_{\alpha=1}^{N}\sum_{i,j=1}^{p} \sigma^{ij\cdot\xi}(x'_{iw\alpha} - \mu_{iw})(x'_{jw\alpha} - \mu_{jw}) + \sum_{w=1}^{2}\sum_{\alpha=1}^{N}\sum_{\xi,\eta=1}^{k} \sigma^{\xi\eta}(x_{\xi w\alpha} - \mu_{\xi w})(x_{\eta w\alpha} - \mu_{\eta w})\right]\right\},$$

where $\sigma^{\xi\eta}$ is the inverse of the $k \times k$ matrix $\sigma_{\xi\eta}$.

We now proceed to estimate Δ^2 in equation (16) by maximum likelihood. For this, we obviously need the sample estimates of the $\sigma^{ij\cdot\xi}$ and the δ_i, and it will appear presently that the sample estimates of the $\beta_{i\xi}$ are also required. However, it happens that the $\sigma^{\xi\eta}$ and the $\mu_{\xi w}$ are not needed. Hence the relevant part of the likelihood function is

$$L = N \log \mid \sigma^{ij\cdot\xi} \mid - \frac{1}{2}\sum_{w=1}^{2}\sum_{\alpha=1}^{N}\sum_{i,j=1}^{p} \sigma^{ij\cdot\xi}(x'_{iw\alpha} - \mu_{iw})(x'_{jw\alpha} - \mu_{jw}) \tag{17}$$

where

$$x'_{iw\alpha} = x_{iw\alpha} - \sum_{\xi=1}^{k} \beta_{i\xi}(x_{\xi w\alpha} - \mu_{\xi w}) .$$

Differentiating first with respect to μ_{iw}, we obtain

$$\sum_{\alpha=1}^{N} \sum_{j=1}^{p} \sigma^{ij \cdot \xi}(x'_{jw\alpha} - \hat{\mu}_{jw}) = 0. \tag{18}$$

Except in the case (with probability zero) where our estimate of $\sigma^{ij \cdot \xi}$ turns out to be singular, these equations have no solution except

$$\sum_{\alpha=1}^{N} (x'_{jw\alpha} - \hat{\mu}_{jw}) = 0 \tag{19}$$

for every j, w. Consequently

$$\hat{\mu}_{jw} = x'_{jw\cdot}$$

so that

$$\hat{\delta}_j = \hat{\mu}_{j2} - \hat{\mu}_{j1} = x'_{j2\cdot} - x'_{j1\cdot} = d_j - \sum_{\xi=1}^{k} \beta_{j\xi}\, d_\xi .$$

This shows that the $\beta_{i\xi}$ must also be estimated. Now

$$\frac{\partial L}{\partial \beta_{i\xi}} = \sum_{w=1}^{2} \sum_{\alpha=1}^{N} \frac{\partial L}{\partial x'_{iw\alpha}} \frac{\partial x'_{iw\alpha}}{\partial \beta_{i\xi}} = \sum_{w=1}^{2} \sum_{\alpha=1}^{N} \sum_{j=1}^{p} \sigma^{ij \cdot \xi}(x_{\xi w\alpha} - \mu_{\xi w})(x'_{jw\alpha} - \mu_{jw}).$$

Once again, unless the estimate of $\sigma^{ij \cdot \xi}$ is singular, the only solutions of the equations formed by equating this quantity to zero are

$$\sum_{w=1}^{2} \sum_{\alpha=1}^{N} (x_{\xi w\alpha} - \mu_{\xi w})(x'_{jw\alpha} - \hat{\mu}_{jw}) = 0 \tag{20}$$

for every ξ, j.

Since $\hat{\mu}_{jw} = x'_{jw\cdot}$, the term in $\mu_{\xi w}$ vanishes. Substituting for x' in terms of x from (17), we obtain

$$\sum_{w=1}^{2} \sum_{\alpha=1}^{N} x_{\xi w\alpha} \left\{ (x_{jw\alpha} - x_{jw\cdot}) - \sum_{\eta=1}^{k} b_{j\eta}(x_{\eta w\alpha} - x_{\eta w\cdot}) \right\} = 0$$

where $b_{j\eta}$ stands for the maximum likelihood estimate of $\beta_{j\eta}$. These equations may be written

$$\sum_{\eta=1}^{k} b_{j\eta} E_{\xi\eta} = E_{j\xi} \tag{21}$$

where E denotes a sum of squares or products of deviations from the sample means, containing $2(N - 1)$ degrees of freedom The equations are therefore

the ordinary normal equations for the 'within-sample' multiple regression of $x_{jw\alpha}$ on the $x_{\xi w\alpha}$.

Finally, differentiation of L with respect to the $\sigma^{ij\cdot\xi}$ leads to

$$2N\hat{\sigma}_{ij\cdot\xi} = \sum_{w=1}^{2}\sum_{\alpha=1}^{N}(x'_{iw\alpha} - x'_{iw\cdot})(x'_{jw\alpha} - x'_{jw\cdot}). \tag{22}$$

This is just the 'within-samples' sum of squares or products of the variates x'. On substituting for the x' in terms of the x and using equations (21), we obtain

$$2N\hat{\sigma}_{ij\cdot\xi} = E_{ij} - \sum_{\xi=1}^{k} b_{i\xi}E_{j\xi} = E_{ij\cdot\xi} \qquad \text{(say)}.$$

To summarize, the estimated distance is given by means of the equation

$$(p + k)D^2 = \sum_{i,j=1}^{p} \hat{\sigma}^{ij\cdot\xi}\hat{\delta}_i\hat{\delta}_j = 2N\sum_{i,j=1}^{p} E^{ij\cdot\xi}\, d'_i\, d'_j,$$

where $E^{ij\cdot\xi}$ is the inverse of $E_{ij\cdot\xi}$ and

$$d'_i = d_i - \sum_{\xi=1}^{k} b_{i\xi}\, d_\xi.$$

This estimate was obtained by assuming *all* variates jointly normally distributed. From the form of the likelihood function (17) it can be seen that the M.L. estimate of the distance remains the same under the less restrictive assumptions that the $x_{\xi w\alpha}$ are fixed, while the deviations of the $x_{iw\alpha}$ from their regressions on the $x_{\xi w\alpha}$ are jointly normal.

13. Computational procedure. An orderly procedure for calculating the generalized distance will now be given. From this, the method for computing the corresponding discriminant function will be shown. The computations also lead to the generalization of Hotelling's T^2. The steps are as follows.

(i). First form the 'within-sample' sums of squares and products of all variates, with $2(N - 1)$ degrees of freedom. These are the quantities denoted by E_{ij}, $E_{i\xi}$, $E_{\xi\eta}$.

(ii). Invert the matrix $E_{\xi\eta}$, giving $E^{\xi\eta}$.

(iii). The regression coefficients $b_{i\xi}$, estimates of the $\beta_{i\xi}$, are now obtainable by means of the relations

$$b_{i\xi} = \sum_{\eta=1}^{k} E_{i\eta}E^{\xi\eta},$$

as is clear from the usual matrix solution of equations (21).

(iv). The sums of squares and products of the deviations of the x_i from their 'within-sample' regressions on the x_ξ are now computed from equations (22)

$$2N\hat{\sigma}_{ij\cdot\xi} = E_{ij\cdot\xi} = E_{ij} - \sum_{\xi=1}^{k} b_{i\xi}E_{j\xi}$$

(v). The final step is to invert the matrix $E_{ij\,\xi}$, giving $E^{ij\,\xi}$, and to form the product

$$(p + k)D^2 = 2N \sum_{i,j=1}^{p} E^{ij\,\xi} d'_i d'_j, \quad \text{where} \quad d'_i = d_i - \sum_{\xi=1}^{k} b_{i\xi} d_\xi .$$

When there were no covariance variates, the discriminant function I had the property that the difference between the two sample means of I was equal to the estimated distance (Section 10). This relationship can be preserved when covariance variates are present by defining I so that

$$I_{w\alpha} = \sum_{i=1}^{p} M_i \left(x_{iw\alpha} - \sum_{\xi=1}^{k} b_{i\xi} x_{\xi w\alpha} \right),$$

and calculating the weights M_i from the equations,

$$\sum_{j=1}^{p} E_{ij\,\xi} M_j = d'_i .$$

For in that case,

$$M_i = \sum_{j=1}^{p} E^{ij\cdot\xi} d'_j .$$

Consequently the difference between the two sample means of I is

$$\sum_{i=1}^{p} M_i d'_i = \sum_{i,j=1}^{p} E^{ij\,\xi} d'_i d'_j ,$$

which (apart from the constant $2N$) is equal to $(p + k)D^2$.

14. Distribution of the estimated distance. In the ordinary case, with no covariance variates, the frequency distribution of the estimated distance has been given by several authors, e.g. Hsu [6]. It will be found that in our problem the distribution is essentially the same, except that the quantity D^2 must be multiplied by a new factor and that one set of degrees of freedom entering into the result must be changed from $(n - p + 1)$ to $(n - p - k + 1)$.

Thus far we have assumed that all variates jointly follow a multivariate normal distribution. It is convenient at this stage to regard the covariance variates $x_{\xi w\alpha}$ as fixed from sample to sample, and to use the conditional distribution of the $x_{iw\alpha}$, subject to this restriction. It is well known (e.g. Cramér [7, section 24.6]) that this conditional distribution is the multivariate normal

$$(23) \qquad \begin{aligned} &(2\pi)^{-Np} \, | \sigma^{ij\cdot\xi} |^{+N} \prod dx_{iw\alpha} \\ &\cdot \exp\left\{ -\tfrac{1}{2} \left[\sum_{w=1}^{2} \sum_{\alpha=1}^{N} \sum_{i,j=1}^{p} \sigma^{ij\,\xi} (x_{iw\alpha} - \mu_{iw} - \gamma_{iw\alpha})(x_{jw\alpha} - \mu_{jw} - \gamma_{jw\alpha}) \right] \right\} \end{aligned}$$

where

$$\gamma_{iw\alpha} = \sum_{\xi=1}^{k} \beta_{i\xi}(x_{\xi w\alpha} - \mu_{\xi w}).$$

Since the estimated distance is a function of the quantities $E_{ij\cdot\xi}$, d'_i, we now find the joint distribution of these variates. The joint distribution of the sums of squares and products $E_{ij\cdot\xi}$ is obtained by quoting a slight extension of a result due to Bartlett [8], which may be stated as follows.

Let the variates $x_{iw\alpha}$ *follow the distribution* (23) *and let*

$$(i) \qquad E_{ij} = \sum_{w=1}^{2} \sum_{\alpha=1}^{N} (x_{iw\alpha} - x_{iw\cdot})(x_{jw\alpha} - x_{jw\cdot})$$

be a typical 'within-samples' sum of squares or products,

$$(ii) \qquad b_{i\xi} = \sum_{\eta=1}^{k} E_{i\eta} E^{\xi\eta}$$

be the 'within-samples' partial regression coefficient of x_i *on* x_ξ, *and*

$$(iii) \qquad E_{ij\cdot\xi} = E_{ij} - \sum_{\xi=1}^{k} b_{i\xi} E_{j\xi}$$

be the sum of squares or products of deviations from these regressions. Then
(*a*) *the quantities* $E_{ij\cdot\xi}$ *follow the Wishart distribution*

$$c \,|\, E_{ij\cdot\xi} \,|^{\frac{1}{2}(n-k-p-1)} \exp\left\{-\tfrac{1}{2} \sum_{i,j=1}^{p} \sigma^{ij\cdot\xi} E_{ij\cdot\xi}\right\} \prod dE_{ij\cdot\xi}$$

with $(n - k)$ *d.f., where* $n = 2(N - 1)$,

(*b*) *this distribution is independent of that of the* $b_{i\xi}$, *and*

(*c*) *both distributions are independent of that of the means* $x_{iw\cdot}$ *and consequently of that of the difference* $d_i = (x_{i2\cdot} - x_{i1\cdot})$.

The result was proved by Bartlett for a sample from a single population. The extension to the case of two populations is straightforward and will not be given in detail.

From (b) and (c) it follows that the distribution of the $E_{ij\cdot\xi}$ is independent of that of the quantities

$$d'_i = d_i - \sum_{\xi=1}^{k} b_{i\xi}\, d_\xi .$$

Further, with the x_ξ variates fixed, the d'_i are linear functions of the $x_{iw\alpha}$ with constant coefficients and hence follow a multivariate normal distribution, Wilks [9]. We now find the means and the covariance matrix of this joint distribution.

From the joint distribution (23) of the $x_{iw\alpha}$, it is easily seen that

$$(24) \qquad E(d_i) = \delta_i + \sum_{\xi=1}^{k} \beta_{i\xi}\, d_\xi .$$

Also, since by standard regression theory the $b_{i\xi}$ are unbiased estimates of the $\beta_{i\xi}$,

$$E\left\{\sum_{\xi=1}^{k} b_{i\xi}\, d_\xi\right\} = \sum_{\xi=1}^{k} \beta_{i\xi}\, d_\xi .$$

Hence, by subtraction,

$$E(d'_i) = \delta_i . \tag{25}$$

Now

$$\text{Cov}\,(d'_i d'_j) = \text{Cov}\,(d_i - \sum_{\xi=1}^{k} b_{i\xi}\, d_\xi)(d_j - \sum_{\eta=1}^{k} b_{j\eta}\, d_\eta).$$

By (c) the distributions of the d_i, $b_{j\eta}$ are independent, so that there will be no contribution from products of the form $d_i b_{j\eta}$. Hence

$$\text{Cov}\,(d'_i\, d'_j) = \text{Cov}\,(d_i\, d_j) + \sum_{\xi,\eta=1}^{k} d_\xi\, d_\eta\, \text{Cov}\,(b_{i\xi} b_{j\eta}). \tag{26}$$

Since d_i is the difference between the means of two samples of size N, Cov $(d_i d_j)$ is $2\,\sigma_{ij\cdot\xi}/N$. The covariance of $b_{i\xi}$ and $b_{j\eta}$ is more troublesome. Writing the expressions for these regression coefficients in terms of the original data, we have

$$\text{Cov}\,(b_{i\xi} b_{j\eta}) = \sum_{\lambda,\nu=1}^{k} E^{\lambda\xi} E^{\nu\eta}\, \text{Cov}\,(E_{i\lambda} E_{j\nu}) =$$

$$\sum_{\lambda,\nu=1}^{k} E^{\lambda\xi} E^{\nu\eta} \sum_{w,z=1}^{2} \sum_{\alpha,\zeta=1}^{N} (x_{\lambda w\alpha} - x_{\lambda w})(x_{\nu z\zeta} - x_{\nu z})\, \text{Cov}\,(x_{iw\alpha} x_{jz\zeta}).$$

Since successive observations are assumed independent, the covariance term vanishes unless $w = z$ and $\alpha = \zeta$, in which case it equals $\sigma_{ij\cdot\xi}$. Thus

$$\text{Cov}\,(b_{i\xi} b_{j\eta}) = \sigma_{ij\cdot\xi} \sum_{\lambda,\nu=1}^{k} E^{\lambda\xi} E^{\nu\eta} E_{\lambda\nu} = \sigma_{ij\cdot\xi} E^{\xi\eta}.$$

Finally, from (26)

$$\text{Cov}\,(d'_i\, d'_j) = \sigma_{ij\cdot\xi}\left(\frac{2}{N} + \sum_{\xi,\eta=1}^{k} E^{\xi\eta}\, d_\xi\, d_\eta\right) = v\sigma_{ij\cdot\xi} \qquad \text{(say)}. \tag{27}$$

Having obtained the distributions of the $E_{ij\cdot\xi}$, d'_i, we may apply Hsu's result [6] for the general distribution of Hotelling's T^2. In our notation, this may be stated as follows.

If the variates $d'_i/\sqrt{v}$ follow the multivariate normal distribution with means $\delta_i/\sqrt{v}$ and covariance matrix $\sigma_{ij\cdot\xi}$, and if the variates $E_{ij\cdot\xi}$ follow the Wishart distribution with $(n - k)$ d.f. and covariance matrix $\sigma_{ij\cdot\xi}$, the two distributions being independent, then

$$y = \sum_{i,j=1}^{p} E^{ij\cdot\xi}\, d'_i\, d'_j/v,$$

follows the distribution

$$e^{-\tau} \sum_{h=0}^{\infty} \frac{\tau^h}{h!} \frac{1}{B\{\frac{1}{2}p + h,\, \frac{1}{2}(n - k - p + 1)\}} \cdot y^{\frac{1}{2}p+h-1}(1 + y)^{-\frac{1}{2}(n-k+1)-h}\, dy, \tag{28}$$

where

$$\tau = \tfrac{1}{2} \sum_{i,j=1}^{p} \sigma^{ij \cdot \xi} \delta_i \delta_j / v,$$

$$v = \frac{2}{N} + \sum_{\xi,\eta=1}^{k} E^{\xi\eta} d_\xi d_\eta, \qquad n = 2(N-1).$$

This distribution is, of course, the distribution of the ratio of two independent values of χ^2, with p and $(n - k - p + 1)$ d.f. respectively, in the case where the numerator is non-central.

15. Tests of significance. This result leads to the extension of Hotelling's T^2 test. For if $\delta_i = 0$, $(i = 1, 2, \cdots p)$, then τ is zero and

$$\sum_{i,j=1}^{p} E^{ij \cdot \xi} d_i' d_j'$$

is distributed as $vpF/(n - k - p + 1)$, with p and $(n - k - p + 1)$ d.f. The distribution (28) above gives the power function of this test.

We may also wish to apply a test of this type to a subgroup x_i of the discriminators $(i = 1, 2, \cdots q < p)$. Speaking popularly, this is a test of the null hypothesis that the above variates x_i contribute nothing to the discrimination between the two populations, given that the remaining discriminators and the covariance variates have already been included.[1] To see what is meant more precisely, consider the following transformation:

$$x_i' = x_i - \Sigma \beta_{il} x_l - \Sigma \beta_{i\xi} x_\xi, \qquad i = 1, 2, \cdots q;$$

$$x_l' = x_l - \Sigma \beta_{l\xi} x_\xi, \qquad l = q + 1, \cdots p;$$

$$x_\xi' = x_\xi, \qquad \xi = 1, 2, \cdots k,$$

where the β's are population regression coefficients. Then it is not difficult to see that the distance is now given by

$$(p + k)\Delta^2 = \sum_{i,j=1}^{q} \sigma'^{ij \cdot l\xi} \delta_i' \delta_j' + \sum_{l,m=q+1}^{p} \sigma^{lm \cdot \xi} \delta_l \delta_m$$

where $\sigma'^{ij \cdot l\xi}$ is the inverse of the covariance matrix of the deviations of the x_i from their regressions on the x_l plus the x_ξ, and

$$\delta_i' = \delta_i - \Sigma \beta_{il} \delta_l .$$

Consequently if $\delta_i' = 0$, $(i = 1, 2, \cdots q)$ the distance is exactly the same as it would be if the variates x_i were omitted. The test in question is therefore a test of the null hypothesis that $\delta_i' = 0$, $(i = 1, 2, \cdots q)$.

If both the remaining discriminators x_l and the covariance variates x_ξ are regarded as fixed, the method of proof in the previous section provides an F test

[1] The test is illustrated in section 7.

for this hypothesis also. It is found that the sums of squares or products $E^{ij \cdot l\xi}$ follow a Wishart distribution with $(n - k - p + q)$ d.f., while the quantities

$$d_i' = d_i - \sum_{l=q+1}^{p} b_{il}\, d_l - \sum_{\xi=1}^{k} b_{i\xi}\, d_\xi$$

are normally distributed, with zero means when the null hypothesis is true. This leads to the result that

$$\sum_{i,j=1}^{q} E^{ij \cdot l\xi}\, d_i'\, d_j'$$

is distributed as $v'qF/(n - k - p + 1)$, with q and $(n - k - p + 1)$ d.f., and

$$v' = \frac{2}{N} + \Sigma\, E^{l\xi}\, d_l\, d_\xi\,,$$

the sum extending over both the covariance variates and the discriminators that are not being tested.

16. Discussion of the gain due to covariance. In this section we attempt to construct a measure of the amount that has been gained by the use of the covariance variates. Only a preliminary discussion will be given: a complete discussion would be rather lengthy, owing to the many different uses to which the discriminant function can be put. Perhaps the problem can most easily be seen by considering the effect on Hotelling's generalized T^2 test of significance.

The power function of this test, as obtained from equation (28) section 14, depends on four factors; the level of significance that is chosen, the degrees of freedom n_1 and n_2 in the numerator and denominator of F, and the parameter τ. If the covariance variates were ignored, the usual T^2 test could be applied to the discriminators alone. In this case we would have

$$n_1' = p, \qquad n_2' = n - p + 1, \qquad \tau' = \tfrac{1}{2}\Sigma\sigma^{ij}\delta_i\delta_j/v', \qquad \text{where } v' = 2/N.$$

With the covariance variates, we have

$$n_1 = p, \qquad n_2 = n - p - k + 1, \qquad \tau = \tfrac{1}{2}\Sigma\sigma^{ij \cdot \xi}\delta_i\delta_j/v,$$

where

$$v = \frac{2}{N} + \Sigma\, E^{\xi\eta}\, d_\xi\, d_\eta\,.$$

The first point to note is that

$$\Sigma\, \sigma^{ij \cdot \xi}\delta_i\delta_j \geq \Sigma\, \sigma^{ij}\delta_i\delta_j$$

This is an instance of the general result that the addition of new variates cannot decrease the value of $p\Delta^2$. To see this, replace the covariance variates by their

deviations from their regressions on the discriminators. This transformation gives

$$\sum_{i,j=1}^{p} \sigma^{ij\cdot\xi}\delta_i\delta_j = \sum_{i,j=1}^{p} \sigma^{ij}\delta_i\delta_j + \sum_{\xi,\eta=1}^{k} \sigma^{\xi\eta\cdot i}\delta'_\xi\delta'_\eta, \tag{29}$$

where

$$\delta'_\xi = \delta_\xi - \sum_{i=1}^{p} \beta_{\xi i}\delta_i .$$

Since the term on the right of equation (29) is a positive definite quadratic form, the result follows.

Consequently, the first effect of the covariance variates is to make the numerator of τ greater than that of τ'. As a partial compensation, the denominator v is also greater than v', but it may be shown that the difference in the denominators will usually be trivial if k is small relative to n. We therefore expect τ to be greater than τ'. Now for fixed n_1, n_2 and significance level, it is well known that the power function (28) is monotone increasing with τ. Hence, other things being equal, the increase in τ due to the covariance variates leads to a more powerful test.

The two power functions, however, differ in another respect, in that with covariance the value of n_2 is reduced from $(n - p + 1)$ to $(n - p - k + 1)$. This decrease in the number of degrees of freedom in the denominator of F will to some extent offset the gain from an increased τ. Examination of Tang's tables [10] indicates, however, that if the degrees of freedom are substantial, this effect will not be important. Moreover, in most practical applications, k is likely to be only 1 or 2. Hence, as a first approximation the effect will be ignored, though to do so tends to overestimate the advantage of covariance.

Suppose now that $\tau = r\tau'$, where $r > 1$. Since τ' is proportional to N, the size of sample taken from each population, we could make $\tau' = \tau$ by increasing the size of sample (when covariance is not used) from N to rN. This suggests that the ratio τ/τ' can be used, as a first approximation, to measure the relative accuracy obtained with and without the use of covariance. This measure carries approximately the usual interpretation that the inferior method would become as good as the superior method if the sample size for the inferior method were increased by the factor r. A further refinement could be made to take account of the difference in the n_2 values. By trial and error applied to Tang's tables, one could determine r' so that the two power functions would be as nearly coincident as possible.

In practice, the ratio τ/τ' must be estimated from the data. From the power function in equation (28) it is found by integration that the mean value of y is

$$(2\tau + p)/(n_2 - 2),$$

so that an unbiased estimate of τ is

$$\tfrac{1}{2}\{(n_2 - 2)y - p\} = \tfrac{1}{2}p\left\{\frac{(n_2 - 2)}{n_2}F - 1\right\}.$$

This suggests that the quantity

$$\frac{(n_2 - 2)}{n_2} F - 1$$

should be calculated both with and without covariance. The ratio of the two values will probably not be an unbiased estimate of τ/τ', but may be used pending further information about its sampling distribution. This type of calculation is made for the numerical example in section 8.

17. The case of a row by column classification. Thus far the discussion has been confined to the case where there are only two populations. The technique may also be used when there are more than two populations. The difference δ_i between the two population means is replaced by some linear function of the population means. As an illustration we consider a row by column classification, the case that arises in the numerical example. No detailed proofs will be given, though it is hoped that the theory can be fairly easily developed from the mathematical model.

A typical variate is x_{iwz}, where $i = 1, 2, \cdots p$ denotes the variate, $w = 1, 2, \cdots r$ denotes the row and $z = 1, 2, \cdots c$ denotes the column, there being one observation in each cell. The variates x_{iwz} follow a multivariate normal distribution, with covariance matrix $\sigma_{ij\xi}$ and means

$$E(x_{iwz}) = \mu_i + \rho_{iw} + \gamma_{iz} + \sum_{\xi=1}^{k} \beta_{i\xi}(x_{\xi wz} - x_{\xi\cdot\cdot}),$$

where ρ_{iw} denotes the effect of the row and γ_{iz} that of the column. Without loss of generality we may assume that

$$\sum_w \rho_{iw} = \sum_z \gamma_{iz} = 0.$$

In addition, there exists a *known* set of variates t_z such that

$$\gamma_{iz} = \delta_i t_z, \qquad \sum_z t_z = 0$$

That is, the column constants have a linear regression on a set of known numbers.

The following are the maximum likelihood estimates of the relevant constants.

$$b_{i\xi} = \sum_{\eta=1}^{k} E_{i\eta} E^{\eta\xi},$$

where

$$E_{i\eta} = \sum_{w,z} x_{iwz} \left\{ x_{\eta wz} - x_{\eta w\cdot} - t_z \frac{(\sum t_z x_{\eta\cdot z})}{\sum t_z^2} \right\}.$$

$$\hat{\delta}_i = \frac{\sum_{w,z} t_z(x_{iwz} - \sum_\xi b_{i\xi} x_{\xi wz})}{\sum t_z^2}.$$

In the notation used for numerical calculation,

$$\hat{\delta}_i = \frac{(d_i - \sum b_{i\xi} d_\xi)}{r \sum t_z^2} = \frac{d_i'}{r \sum t_z^2}, \qquad \text{where} \quad d_i = \sum_z t_z X_{i \cdot z},$$

the quantity $X_{i\,z}$ being the column *total* Finally

$$rc\hat{\sigma}_{ij\,\xi} = E_{ij\,\xi} = E_{ij} - \sum_\xi b_{i\xi} E^{\xi j}.$$

The distributional properties are similar to those in the two-population case. The quantities $E_{ij\cdot\xi}$ follow a Wishart distribution ith $(rc - r - 1 - k)$ d.f. and covariance matrix $\sigma_{ij\,\xi}$. The variates d_i' follow a multivariate normal distribution with means $r\delta_i \Sigma t_z^2$ and covariance

$$\sigma_{ij\,\xi}(r\Sigma t_z^2 + \Sigma E^{\xi\eta} d_\xi d_\eta) = v\sigma_{ij\cdot\xi} \qquad \text{(say)}.$$

Consequently,

$$y = \Sigma E^{ij\cdot\xi} d_i' d_j'$$

is distributed as $vpF/(rc - r - p - k)$ with p and $(rc - r - p - k)$ d.f. and parameter

$$\tau = \tfrac{1}{2}(r\Sigma t_z^2)^2 \Sigma \sigma^{ij\cdot\xi} \delta_i \delta_j / v.$$

Thus in the numerical example, with $r = 12$, $c = 4$, $p = 2$, $k = 1$, this procedure would have given an F test of the null hypothesis $\tau = 0$, where F has 2 and 33 d.f. However, the contribution from 2 degrees of freedom was deliberately omitted from the quantities E_{ij}, so that F actually had 2 and 31 d.f.

PART III

18. Justification of the 'dummy variate' approach. It remains to show that the method of calculation used in the example (sections 5 and 6) is equivalent to that derived from theory. There are two chief points to prove. First, that the M's found from the equations

(30) $$\sum_j E_{ij\cdot\xi} M_j = d_i'$$

are proportional to the corresponding L's found from the equations

(31) $$\sum_a S_{ij} L_j = d_i$$

where the suffix a denotes summation over both x_i and x_ξ variates.

Now, since $S_{ij} = E_{ij} + d_i d_j/240$, equations (31) are the same as

(32) $$\sum_a E_{ij} L_j = d_i(1 - \sum_a L_j d_j/240).$$

Hence the L's in (31) are proportional to the values found from the equations

(33) $$\sum_a E_{ij} L_j' = d_i\,.$$

But it is well known that if the L'_ξ are eliminated one by one from equations (33), we obtain

$$\sum_j E_{ij \cdot \xi} L'_j = d'_i ,$$

which is the same as (30). This proves the first point.

The second point to establish is that the F test in the example is the same as that obtained from theory. In section 15, it was shown that

$$\sum_{i,j} E^{ij \cdot \xi} d'_i d'_j / v \tag{34}$$

is distributed as $pF/(n - p - k + 1)$. In the analysis of variance of Table 6, section 6, the quantity follo ring the same distribution was

$$\frac{(S_a - S_\xi)}{(240 - S_a)}, \tag{35}$$

where

$$S_a = \sum_a S^{ij} d_i d_j , \qquad S_\xi = \sum_{\xi,\eta} S^{\xi\eta} d_\xi d_\eta$$

Since equations (31) and (32) have the same solution, we must have

$$S^{ij} = E^{ij}(1 - \sum_a L_j d_j/240) = E^{ij}(1 - S_a/240).$$

Multiplying both sides by $d_i d_j$ and summing over all i, j, we obtain

$$S_a = E_a(1 - S_a/240) = E_a(1 + E_a/240),$$

where E_a is defined analogously to S_a. Similarly

$$S_\xi = E_\xi/(1 + E_\xi/240).$$

Hence

$$\frac{S_a - S_\xi}{240 - S_a} = \frac{E_a - E_\xi}{240 + E_\xi} = \frac{E_a - E_\xi}{v}. \tag{36}$$

Transform the variates x_i, x_ξ into variates x'_i, x_ξ, where $x'_i = x_i - \Sigma b_{i\xi} x_\xi$. It is easy to see that this transforms

$$\sum_a E^{ij} d_i d_j \quad \text{into} \quad \sum E^{\xi\eta} d_\xi d_\eta + \sum_{i,j} E^{ij \cdot \xi} d'_i d'_j .$$

That is,

$$E_a = E_\xi + \sum_{i,j} E^{ij \cdot \xi} d'_i d'_j ,$$

since the quantity on the left is invariant under non-singular linear transformations. Hence from (36),

$$\frac{(S_a - S_\xi)}{(240 - S_a)} = \sum_{i,j} E^{ij \cdot \xi} d'_i d'_j / v.$$

From (34) and (35), this establishes the equivalence of the F tests. While the proof has been given only for the type of data encountered in the example, the same method will apply to other types of data.

In conclusion, we wish to thank the referees for many helpful suggestions in connection with the presentation of this paper.

REFERENCES

[1] R. A. FISHER, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, Vol. 7 (1936), pp. 179–188

[2] R. A. FISHER, "The statistical utilization of multiple measurements," *Annals of Eugenics*, Vol 8 (1938), pp. 376–386.

[3] H HOTELLING, "The generalization of Student's ratio," *Annals of Math. Stat.*, Vol. 2 (1931), pp. 360–378.

[4] P. C. MAHALANOBIS, "On the generalized distance in statistics," *Proc. Nat. Inst. Sci Ind.*, Vol. 12 (1936), pp 49–55

[5] C I BLISS AND H. P. MARKS, "The biological assay of insulin," *Quart. Jour. Pharm. and Pharmacol*, Vol 12 (1939), pp. 82–110; 182–205.

[6] P. L. HSU, "Notes on Hotelling's generalized T," *Annals of Math. Stat.*, Vol. 9 (1938), pp 231–243

[7] H. CRAMÉR, *Mathematical methods of statistics*, Princeton University Press, 1946.

[8] M S BARTLETT, "On the theory of statistical regression," *Roy. Soc. Proc. Edin.*, Vol. 53 (1933), pp. 271–277

[9] S S. WILKS, *Mathematical Statistics*, Princeton University Press, 1943, p. 70.

[10] P. C. TANG, "The power function of the analysis of variance tests," *Stat. Res. Memoirs*, Vol. 2 (1938), pp 126–157.

ON THE KOLMOGOROV-SMIRNOV LIMIT THEOREMS FOR EMPIRICAL DISTRIBUTIONS

By W. Feller

Cornell University[*]

Summary. Unified and simplified derivations are given for the limiting forms of the difference (1) between the empirical distribution of a large sample and the corresponding theoretical distribution and (2) between the distributions of two large samples.

1. Introduction. Let $X_1, \cdots, X_N$ be mutually independent random variables with the common cumulative distribution function $F(x)$. Let $X_1^*, \cdots, X_N^*$ be the same set of variables rearranged in increasing order of magnitude. *The empirical distribution* (or sum-polygon) *of the sample* $X_1, \cdots, X_N$ *is the step function* $S_N(x)$ *defined by*

$$S_N(x) = \begin{cases} 0 & \text{for } x < X_1^* \\ \dfrac{k}{N} & \text{for } X_k^* \le x < X_{k+1}^* \\ 1 & \text{for } x \ge X_N^*. \end{cases} \tag{1.1}$$

In other words, $N \cdot S_N(x)$ equals the number of variables X_ν which do not exceed x. We expect intuitively that $S_N(x) \to F(x)$ as $N \to \infty$. In fact, if this were not so the notions of distribution and sample would be meaningless. The so-called ω^2-criterion of von Mises [4] provides rough estimates for the probable deviations of $S_N(x)$ from $F(x)$ for certain forms of $F(x)$ (cf. von Mises [4]). A much stronger result is due to A. Kolmogorov and is of great interest in the theory of non-parametric estimation (Kolmogorov [3]). The maximum of the deviation $| S_N(x) - F(x) |$ is a random variable D_N whose distribution is easily seen to be independent of the special form of $F(x)$ provided only that $F(x)$ is continuous.[1] The exact distribution of D_N is not known, but Kolmogorov found that $N^{\frac{1}{2}} D_N$ has a limiting distribution. More precisely we have

Theorem 1 (Kolmogorov [1]). *Suppose that* $F(x)$ *is continuous and define the random variable* D_N *by*

$$D_N = \text{l.u.b.} | S_N(x) - F(x) |. \tag{1.2}$$

*Research under an ONR contract.

[1] This fact will not be used explicitly in the sequel but follows as a byproduct from our proofs. A simple direct proof consists in considering the random variables $\Xi_k = F(X_k)$ which are uniformly distributed, the maximum deviation D_N of the empirical distribution of the new sample $\{\Xi_k\}$ from the uniform distribution has the same distribution as D_N; cf. Kolmogorov [1].

Then for every fixed $z \geq 0$ as $N \to \infty$

$$\Pr \{D_N \leq zN^{-\frac{1}{2}}\} \to L(z) \tag{1.3}$$

where $L(z)$ is the cumulative distribution function which for $z > 0$ is given by either of the following equivalent relations[2]

$$L(z) = 1 - 2 \sum_{\nu=1}^{\infty} (-1)^{\nu-1} e^{-2\nu^2 z^2} = (2\pi)^{\frac{1}{2}} z^{-1} \sum_{\nu=1}^{\infty} e^{-(2\nu-1)^2 \pi^2 / 8z^2} \tag{1.4}$$

For $z \leq 0$ we have, of course, $L(z) = 0$.

Equally interesting is Smirnov's result concerning the maximum difference between the empirical distributions of two samples with the same cumulative distribution.

THEOREM 2 (Smirnov [5]). *Let $(X_1, \cdots, X_m)$ and $(Y_1, \cdots, Y_n)$ be two samples of mutually independent random variables having a common continuous distribution $F(x)$. Let $S_m(x)$ and $T_n(x)$ be the corresponding empirical distribution functions and define a new random variable $D_{m,n}$ by*

$$D_{m,n} = \text{l.u.b.} | S_m(x) - T_n(x) |. \tag{1.5}$$

Put

$$N = \frac{mn}{m+n} \tag{1.6}$$

and suppose that $m \to \infty$, $n \to \infty$ so that

$$\frac{m}{n} \to a, \tag{1.7}$$

where a is a constant. Then for every fixed $z \geq 0$

$$\Pr \{D_{m,n} \leq zN^{-\frac{1}{2}}\} \to L(z), \tag{1.8}$$

where $L(z)$ is the same as in (1.4).

The original proofs (Kolmogorov [1] and Smirnov [6]) are very intricate and are based on completely different methods. Kolmogorov's proof is based on an auxiliary theorem of equal depth proved in a separate paper (Kolmogorov [2]). An alternative proof of Kolmogorov's theorem is due to Smirnov [5]. However, Smirnov derives both theorems as corollaries to much deeper (but less useful) results concerning the number of intersections of the graphs of $S_N(x)$ and $F(x) \pm \epsilon N^{-\frac{1}{2}}$ and of $S_m(x)$ and $T_n(x) \pm \epsilon N^{-\frac{1}{2}}$, respectively. It is, therefore, not surprising that Smirnov's proofs require a powerful technique and many auxiliary considerations. It is the purpose of the present paper to present unified proofs of the two theorems which are based on methods of great generality.[3] The new

[2] The equivalence of the two formulas in (1.4) is a well-known relation often called transformation formula for theta-functions. We shall only prove the first representation in (1.4). The second is more useful for small z. A table of $L(z)$ is given in Smirnov [6]. It is reprinted in the present issue of the *Annals of Mathematical Statistics* (pp. 279–281).

[3] Among other results which can be proved by the same method are certain limit theorems for ruin and first-passage time problems in the theory of diffusion and random walks.

proof is not simple but simpler than the original ones. At any rate, it requires essentially only routine manipulations with generating functions and their limiting form, the Laplace transforms. However, the paper aims mostly at a unification of methods.

As a byproduct of the proof we obtain

THEOREM 3. *Let A_N be the number of points x where the step-polygon $S_N(x)$ of Theorem 1 leaves the strip $F(x) \pm zN^{-\frac{1}{2}}$. The expected value of the random variable A_N satisfies the asymptotic relation*

$$E(A_N) \sim 2(2\pi N)^{\frac{1}{2}}\{1 - \Phi(2z)\}, \tag{1.9}$$

where $\Phi(z)$ is the normalized Gaussian distribution.

An analogous corollary to Theorem 2 was given by Smirnov [8]: formula (1.9) holds also for the number of intersections of the graph of $S_m(x)$ with the step-polygons $T_n(x) \pm zN^{-\frac{1}{2}}$. These results should come as a surprise to most statisticians. According to Theorem 1 there is a positive probability that $A_N = 0$ and nevertheless $E(A_N)$ is of the order of magnitude $N^{\frac{1}{2}}$. The explanation lies in the fact that *if* $S_N(x)$ crosses the curve $F(x) + zN^{-\frac{1}{2}}$ at some point then it is extremely likely that $S(x)$ will in some neighborhood continue to fluctuate around values $F(x) + zN^{-\frac{1}{2}}$, crossing that curve a great many times. The difference $S_N(x) - F(x)$ exhibits, in the limit $N \to \infty$, many small oscillations. This phenomenon is related to the well-known fact that the path of a particle subject to the Einstein-Wiener diffusion process has no derivatives.

Instead of the absolute values of the differences we may consider the differences themselves and derive two parallel theorems for the maximum and the minimum. As an example we shall prove

THEOREM 4. *With the notations and assumptions of Theorem 1 let*

$$D_N^+ = \text{l.u.b.}\{S_N(x) - F(x)\}. \tag{1.10}$$

Then

$$\Pr\{D_N^+ \le zN^{-\frac{1}{2}}\} \to 1 - e^{-2z^2}. \tag{1.11}$$

The proof is simpler than that of Theorem 1 but uses the same method.

2. Notations and preliminary remarks. For printing convenience it is desirable to avoid complicated subscripts and we shall therefore use the following notation for *binomial coefficients*

$$C(n, k) = \binom{n}{k}. \tag{2.1}$$

Similarly, for the general term of the binomial distribution we shall write

$$B(n, k; p) = C(n, k)p^k(1 - p)^{n-k}. \tag{2.2}$$

If A is an event, $\bar{A}$ will denote its negation (complementary event). Finally

$$\Pr\{A \mid B\} \tag{2.3}$$

denotes the conditional probability of A for given B.

Our proofs depend on a special case of the continuity theorem for characteristic functions. Since we shall deal only with probability density functions $f(t)$ which vanish for $t < 0$ it is preferable to use, instead of the characteristic function, the *Laplace transform*

$$\phi(s) = \int_0^\infty e^{-st} f(t)\, dt. \tag{2.4}$$

(This amounts to using the variable $-s$ instead of the usual is and therefore $\phi(s)$ obeys the formal rules for characteristic functions.)

For any sequence $\{u_k\}(k = 1, 2, \cdots)$ of non-negative numbers we define the *generating function* $u(\lambda)$ *by*

$$u(\lambda) = \sum_{k=1}^{\infty} u_k \lambda^k. \tag{2.5}$$

Now let $\delta > 0$ be fixed and consider the step-function $f_\delta(t)$ defined by

$$f_\delta(t) = u_k \quad \text{for } (k-1)\delta \leq t < k\delta \tag{2.6}$$

($k = 1, 2, \cdots; f_\delta(t) = 0$ for $t < 0$). Its Laplace transform is

$$\phi_\delta(s) = \frac{e^{\delta s} - 1}{s} u(e^{-\delta s}). \tag{2.7}$$

We have, therefore, the *continuity-theorem: If, as* $\delta \rightarrow 0$,

$$\delta u(e^{-\delta s}) \rightarrow \phi(s), \tag{2.8}$$

then for every fixed $t > 0$

$$u_k \rightarrow f(t) \quad \textit{when } k\delta \rightarrow t; \tag{2.9}$$

conversely, if (2.9) *holds then* (2.8) *is true.*

3. Proof of Theorem 1. Since $F(x)$ is continuous it is possible to define numbers x_k such that

$$F(x_k) = \frac{k}{N}, \qquad (k = 1, 2, \cdots, N-1). \tag{3.1}$$

This definition is unique except when $F(x) = k/N$ within an entire interval, in which case we define x_k as the *left* endpoint of that interval.

Let $c > 0$ be an integer. We shall evaluate the probability of the event $D_N > c/N$ and we shall later put

$$c = zN^{\frac{1}{2}}, \qquad N \rightarrow \infty. \tag{3.2}$$

Suppose first that for some particular x

$$S_N(x) - F(x) > \frac{c}{N}. \tag{3.3}$$

This point x is contained in a maximal interval in which (3.3) holds and at the right endpoint ξ of this interval we shall have

$$S_N(\xi) - F(\xi) = \frac{c}{N}. \tag{3.4}$$

Now $S_N(\xi)$ is necessarily a number of the form r/N with an integer r. Since c is an integer also $F(\xi) = k/N$ and hence $\xi = x_k$ for some k. From (3.4) we conclude that

$$X^*_{k+c} < x_k, \qquad X^*_{k+c+1} > x_k \tag{3.5}$$

or in other words: exactly $k + c$ among the N variables X_ν are smaller than x_k. Denote this event by $A_k(c)$. The inequality (3.3) takes place for some x if, and only if, at least one among the events $A_1(c), \cdots, A_N(c)$ occurs. The argument applies equally to $c < 0$ and shows that *the event $D_N > c/N$ occurs if, and only if, at least one among the events*

$$A_1(c),\ A_1(-c),\ A_2(c),\ A_2(-c), \cdots,\ A_N(c),\ A_N(-c) \tag{3.6}$$

occurs.

Let U_r and V_r be the events that in the sequence (3.6) the first event to occur are $A_r(c)$ or $A_r(-c)$, respectively. More formally, the events U_r and V_r are defined by

$$\begin{aligned} U_r &= \bar{A}_1(c)\bar{A}_1(-c) \cdots \bar{A}_{r-1}(c)\bar{A}_{r-1}(-c)A_r(c) \\ V_r &= \bar{A}_1(c)\bar{A}_1(-c) \cdots \bar{A}_{r-1}(c)\bar{A}_{r-1}(-c)\bar{A}_r(c)A_r(-c). \end{aligned} \tag{3.7}$$

These events are mutually exclusive and therefore

$$\Pr\left\{ D_N > \frac{c}{N} \right\} = \sum_{r=1}^{N} \Pr\{U_r\} + \sum_{r=1}^{N} \Pr\{V_r\}. \tag{3.8}$$

From the very definitions we have the following two *fundamental relations*

$$\begin{aligned} \Pr\{A_k(c)\} &= \sum_{r=1}^{k} \Pr\{U_r\} \Pr\{A_k(c) \mid A_r(c)\} \\ &\qquad + \sum_{r=1}^{k} \Pr\{V_r\} \Pr\{A_k(c) \mid A_r(-c)\} \\ \Pr\{A_k(-c)\} &= \sum_{r=1}^{k} \Pr\{U_r\} \Pr\{A_k(-c) \mid A_r(c)\} \\ &\qquad + \sum_{r=1}^{k} \Pr\{V_r\} \Pr\{A_k(-c) \mid A_r(-c)\}. \end{aligned} \tag{3.9}$$

This is a system of $2N$ linear equations for the $2N$ unknowns $\Pr\{U_r\}$ and $\Pr\{V_r\}$ and we proceed to solve it by the method of generating functions.

By definition of x_k we have $\Pr\{X_\nu < x_k\} = k/N$. The probability of the event $A_k(c)$ (that the same inequality holds for exactly $k + c$ different ν's) is therefore given by

$$\Pr\{A_k(c)\} = B(N, k + c; k/N) \tag{3.10}$$

(cf. (2.2)). Similarly, it is readily verified that for $r \leq k$

$$\Pr\{A_k(c) \mid A_r(c)\} = B(N - r - c, k - r; (k - r)/(N - r)). \tag{3.11}$$

and

$$\Pr\{A_k(c) \mid A_r(-c)\} = B(N - r + c, k - r + 2c; (k - r)/(N - r)). \tag{3.12}$$

The last three equations hold also for $c < 0$. They can be written in a more convenient form in terms of the quantities

$$p_k(c) = e^{-k} \frac{k^{k+c}}{(k + c)!}. \tag{3.13}$$

In fact

$$\Pr\{A_k(c)\} = \frac{p_k(c)p_{N-k}(-c)}{p_N(0)} \tag{3.14}$$

$$\Pr\{A_k(c) \mid A_r(c)\} = \frac{p_{k-r}(0)p_{N-k}(-c)}{p_{N-r}(-c)} \tag{3.15}$$

$$\Pr\{A_k(c) \mid A_r(-c)\} = \frac{p_{k-r}(2c)p_{N-k}(-c)}{p_{N-r}(c)} \tag{3.16}$$

If these expressions are introduced into (3.9) the second factor in the numerator cancels. A further simplification is achieved on introducing new sets of unknowns

$$u_r = \Pr\{U_r\} \frac{p_N(0)}{p_{N-r}(-c)} \qquad v_r = \Pr\{V_r\} \frac{p_N(0)}{p_{N-r}(c)}. \tag{3.17}$$

The fundamental equations (3.6) then reduce to

$$\begin{aligned} p_k(c) &= \sum_{r=1}^{k} u_r p_{k-r}(0) + \sum_{r=1}^{k} v_r p_{k-r}(2c) \\ p_k(-c) &= \sum_{r=1}^{k} u_r p_{k-r}(-2c) + \sum_{r=1}^{k} v_r p_{k-r}(0). \end{aligned} \tag{3.18}$$

This system is of the convolution type and can therefore be solved by means of generating functions. The essential point is that the $p_k(c)$ are defined for all k and that the system (3.18) therefore determines the unknowns u_r and v_r for all $r > 0$. We put

$$u(\lambda) = \sum_{k=1}^{\infty} u_k \lambda^k \qquad v(\lambda) = \sum_{k=1}^{\infty} v_k \lambda^k \tag{3.19}$$

and

$$p(\lambda; c) = N^{-\frac{1}{2}} \sum_{k=1}^{\infty} p_k(c)\, \lambda^k . \tag{3.20}$$

(The factor $N^{-\frac{1}{2}}$ serves to simplify formulas.) Then obviously

$$\begin{aligned} p(\lambda; c) &= u(\lambda)p(\lambda; 0) + v(\lambda)p(\lambda; 2c); \\ p(\lambda; -c) &= u(\lambda)p(\lambda; -2c) + v(\lambda)p(\lambda; 0). \end{aligned} \tag{3.21}$$

From here we find $u(\lambda)$ and $v(\lambda)$. Equation (3.17) then determines Pr $\{U_r\}$ and Pr $\{V_r\}$. Actually we are interested only in the two sums occurring in (3.8). We put

$$\xi_k = \frac{1}{p_N(0)} \sum_{r=1}^{k} p_{k-r}(-c)u_r , \qquad \eta_k = \frac{1}{p_N(0)} \sum_{r=1}^{k} p_{k-r}(c)v_r . \tag{3.22}$$

Again, the ξ_k and η_k are defined for all k (also $k \geq N$). From (3.17) we have

$$\sum_{r=1}^{N} \Pr \{U_r\} = \xi_N , \qquad \sum_{r=1}^{N} \Pr \{V_r\} = \eta_N , \tag{3.23}$$

and hence finally, by (3.8)

$$\Pr \{D_N > c/N\} = \xi_N + \eta_N . \tag{3.24}$$

In (3.22) we find again simple convolutions leading to products of the corresponding generating functions. Thus

$$\begin{aligned} \xi(\lambda) &= \sum_{k=1}^{\infty} \xi_k \lambda^k = \frac{u(\lambda)p(\lambda; -c)N^{\frac{1}{2}}}{p_N(0)} \\ \eta(\lambda) &= \sum_{k=1}^{\infty} \eta_k \lambda^k = \frac{v(\lambda)p(\lambda; c)N^{\frac{1}{2}}}{p_N(0)} . \end{aligned} \tag{3.25}$$

We now pass to a study of the limiting form of these generating functions as $N \to \infty$ and $c \to \infty$ in accordance with (3.2). Consider a fixed $t > 0$ and suppose that

$$\frac{k}{N} \to t. \tag{3.26}$$

From well-known properties of the Poisson distribution it follows then that

$$N^{\frac{1}{2}}p_k(c) \to (2\pi t)^{-\frac{1}{2}}\exp(- z^2/2t). \tag{3.27}$$

Accordingly, the continuity theorem of section 2 implies (as can be verified directly) that

$$\begin{aligned} p(e^{-s/N} ; zN^{\frac{1}{2}}) &\to (2\pi)^{-\frac{1}{2}} \int_0^{\infty} t^{-\frac{1}{2}} \exp\,(-ts - z^2/2t)\, dt \\ &= (2s)^{-\frac{1}{2}} \exp(-(2sz^2)^{\frac{1}{2}}). \end{aligned} \tag{3.28}$$

(the last integral is well known and can be evaluated by elementary methods; the square-root is always positive). We see in particular that the limiting form is the same for $p(\lambda; c)$ and $p(\lambda; -c)$. It follows therefore from (3.21) directly that

$$\lim_{N\to\infty} u(e^{-s/N}) = \lim_{N\to\infty} v(e^{-s/N}) = \frac{\exp(-(2sz^2)^{\frac{1}{2}})}{1+\exp(-(8sz^2)^{\frac{1}{2}})}. \tag{3.29}$$

Using this and the fact that $p_N(0) \to (2\pi N)^{-\frac{1}{2}}$ we conclude from (3.25) that

$$\begin{aligned}\lim_{N\to\infty} N^{-1}\xi(e^{-s/N}) &= \lim_{N\to\infty} N^{-1}\eta(e^{-s/N}) \\ &= \left(\frac{2\pi}{2s}\right)^{1/2} \frac{\exp(-(8sz^2)^{1/2})}{1+\exp(-(8sz^2)^{1/2})} = \phi(s).\end{aligned} \tag{3.30}$$

Expanding $\phi(s)$ into a geometric series we get

$$\phi(s) = \left(\frac{2\pi}{2s}\right)^{1/2} \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \exp(-(8s\nu^2 z^2)^{1/2}). \tag{3.31}$$

From the evaluation of the integral in (3.28) we conclude that $\phi(s)$ is the Laplace transform of

$$f(t) = \sum_{\nu=1}^{\infty} (-1)^{\nu-1} \exp(-2\nu^2 z^2/t). \tag{3.32}$$

The continuity theorem of section 2 in conjunction with (3.30) and (3.26) shows that

$$\lim_{N\to\infty} \xi_N = \lim_{N\to\infty} \eta_N = f(1). \tag{3.33}$$

In view of (3.24) this accomplishes the proof.

4. Proof of Theorem 4. This proof is simpler than the preceding one inasmuch as we are now interested only in the events $A_k(c)$ for $c > 0$. This time we define U_r as the event that k is the smallest subscript for which $A_k(c)$ occurs, that is, $U_r = \bar{A}_1(c)\bar{A}_2(c) \cdots \bar{A}_{r-1}(c)A_r(c)$; no analogue to the event V_r will be used. With the same notations as before (3.9) is replaced by

$$\Pr\{A_k(c)\} = \sum_{r=1}^{k} \Pr\{U_r\} \Pr\{A_k(c) \mid A_r(c)\}, \tag{4.1}$$

and hence (3.21) by

$$p(\lambda; c) = u(\lambda)p(\lambda; 0). \tag{4.2}$$

Here $p(\lambda; c)$ is the same as before, so that (cf. (3.29))

$$\lim_{N\to\infty} u(e^{-s/N}) = \exp(-(2sz^2)^{1/2}). \tag{4.3}$$

Again, the first equation (3.25) holds without change and therefore we get instead of (3.30)

$$\lim_{N\to\infty} N^{-1}\xi(e^{-s/N}) = \left(\frac{2\pi}{2s}\right)^{1/2} \exp(-(8sz^2)^{1/2}) \tag{4.4}$$

From (3.28) this is the Laplace transform of

(4.5) $$f(t) = t^{-\frac{1}{2}} \exp(-2z^2/t).$$

As before we conclude that $\zeta_N \to f(1)$, which accomplishes the proof.

5. Proof of Theorem 3. We have seen in section 3 that the intervals in which (3.3) holds are in a one-to-one correspondence with the events $A_k(c)$. Hence

(5.1) $$E(A_N) = \Sigma \Pr\{A_k(c)\} + \Sigma \Pr\{A_k(-c)\}.$$

To evaluate the sums we use (3.10). If $N \to \infty$ and again $c = zN^{\frac{1}{2}}$, $k/N \to t$, then by the central limit theorem

(5.2) $$B(N, k + c; k/N) \to \frac{\exp(-z^2/2t(1-t))}{(2\pi Nt(1-t))^{1/2}}.$$

It follows then from (3.10) that

(5.3) $$N^{-1/2}\Sigma \Pr\{A_k(c)\} \to (2\pi)^{-1/2} \int_0^1 \{t(1-t)\}^{-1/2} \exp(-z^2/2t(1-t))\, dt.$$

Call the right hand member R. After the substitution $t = \sin^2(\phi/2)$ we find

$$\frac{dR}{dz} = -8(2\pi)^{-1/2} z \int_0^{\pi/2} \sin^{-2}\phi \exp(-2z^2/\sin^2\phi)\, d\phi$$

(5.4) $$= 8(2\pi)^{-1/2} z \exp(-2z^2) \int_0^{\pi/2} \exp(-2z^2 \cot^2\phi)\, d(\cot\phi)$$

$$= -2\exp(-2z^2).$$

Since $R \to 0$ as $z \to \infty$ we conclude that

(5.5) $$R = 2\int_z^{\infty} \exp(-2x^2)\, dx = \{1 - \Phi(2z)\}(2\pi)^{1/2}.$$

The same asymptotic estimate holds for the other sum in (5.1), and hence Theorem 3 is proved.

6. Proof of Theorem 2. Reorder the two samples in ascending order of magnitude and denote the rearranged samples by $(X_1^*, \cdots, X_m^*)$ and $(Y_1^*, \cdots, Y_n^*)$. When speaking of the graphs of the empirical distributions $S_m(x)$ and $T_n(x)$ we shall suppose that they have been completed by adding vertical segments so that the graphs become step-polygons. We shall put

(6.1) $$\frac{m}{m+n} = p, \qquad \frac{n}{m+n} = q.$$

Then, according to (1.6) and (1.7)

(6.2) $$\frac{p}{q} \to a, \qquad N = pn = qm.$$

Without loss of generality we shall suppose that

(6.3) $$p \leqq q.$$

In order to carry over the proof of Theorem 1 it is necessary to define the events $A_k(c)$ in a judicious manner. For every integer $k > 0$ let ν_k be the number of variables X_ν which are smaller than Y_k. In other words, ν_k is defined as the integer for which

(6.4) $$X^*_{\nu_k} < Y^*_k \leq X^*_{\nu_k+1}.$$

Finally put

(6.5) $$a_k = \left[\frac{mk}{n}\right] = \left[\frac{p}{q}\, k\right]$$

where, as usual, $[x]$ denotes the greatest integer contained in x.

For $0 < k \leq n$ let $A_k(c)$ be the event that

(6.6) $$\nu_k = a_{k+c}.$$

The possibility of applying the proof of section 1 depends on the following

LEMMA. *Whenever*

(6.7) $$D_{m,n} > \frac{c}{n} > 0$$

then at least one among the events $A_1(c)$, $A_1(-c)$, $\cdots$, $A_n(c)$, $A_n(-c)$ *occurs. Conversely, if one of these events occurs then*

(6.8) $$D_{m,n} > \left(c - \frac{q}{p}\right) \Big/ n.$$

PROOF. If (6.7) holds then either for some x_0

(6.9) $$S_m(x_0) - T_n(x_0) > \frac{c}{n}$$

or the reversed inequality holds with c replaced by $-c$. It suffices to consider the case (6.9). For sufficiently large x we have $S_m(x) = T_n(x) = 1$. Hence the graphs of $S_m(x)$ and $T_n(x) + c/n$ must intersect at an abscissa $\xi > x_0$. The point of intersection lies necessarily on a horizontal segment of the graph of $S_m(x)$ and a vertical segment of $T_n(x) + c/n$. Hence there exists a k such that $\xi = Y_k^*$ and, moreover,

(6.10) $$T_n(\xi -) + \frac{c}{n} < S_m(\xi) \leq T_n(\xi +) + \frac{c}{n}.$$

This amounts to saying that

(6.11) $$\frac{k - 1 + c}{n} < \frac{\nu_k}{m} < \frac{k + c}{n}.$$

In view of (6.3) and (6.5) this relation implies (6.6).

Conversely, suppose that the event $A_k(c)$ occurs and let $c > 0$. Put again $\xi = Y_k^*$. Then, by definition,

$$S_m(\xi) = \frac{\nu_k}{m} = \frac{a_{k+c}}{m}, \qquad T_n(\xi) = \frac{k}{n}. \tag{6.12}$$

It follows that

$$S_m(\xi) > \frac{k+c}{n} - \frac{1}{m} = T_n(\xi) + \frac{c}{n} - \frac{1}{m}, \tag{6.13}$$

which in turn implies (6.8). This proves the lemma.

Theorem 2 is concerned with values of c such that $cn^{-1} = zN^{-\frac{1}{2}}$; in passing to the limit we must therefore put

$$c = z(n/p)^{\frac{1}{2}}. \tag{6.14}$$

Accordingly, the relations (6 7) and (6 8) are asymptotically equivalent and our lemma shows that, asymptotically, the probability of (6.7) is the same as the probability that at least one among the events $A_1(c), \cdots, A_N(-c)$ occurs. To evaluate this probability we proceed exactly as in section 3. The events U_r and V_r defined by (3.7) and the fundamental relations (3.9) hold again. However (3.10) — (3.12) have to be replaced by new evaluations.

It is easily seen that the probability that exactly r among the X_ν are smaller than Y_k^* is the same as the probability to extract exactly r white balls before the k-th black ball from an urn containing m white and n black balls (assuming that all orders are equally likely and that balls are not replaced). In this way one finds

$$\Pr\{A_k(c)\} = \frac{C(a_{k+c}+k-1, k-1)C(m+n-a_{k+c}-k, n-k)}{C(m+n, n)} \tag{6.15}$$

$$\Pr\{A_k(c) \mid A_r(c)\} = \frac{C(a_{k+c}-a_{r+c}+k-r-1, k-r-1)C(m+n-a_{k+c}-k, n-k)}{C(m+n-a_{r+c}-r, n-r)} \tag{6.16}$$

$$\Pr\{A_k(c) \mid A_r(-c)\} = \frac{C(a_{k+c}-a_{r-c}+k-r-1, k-r-1)C(m+n-a_{k+c}-k, n-k)}{C(m+n-a_{r-c}-r, n-r)}. \tag{6.17}$$

The second binomial coefficient in the numerator is common to the three expressions and cancels when the expressions are introduced into (3.9). These fundamental relations assume a more natural form if the occurring binomial coefficients are enlarged to terms of a binomial distribution. It is easily verified that the first of the equations (3.9) reduces to

$$\begin{aligned} &\frac{B(a_{k+c}+k-1, k-1; q)}{B(m+n, n; q)} \\ &= \sum_{r=1}^{k} \Pr\{U_r\} \frac{B(a_{k+c}-a_{r+c}+k-r-1, k-r-1; q)}{B(m+n-a_{r+c}-r, n-r; q)} \\ &\qquad + \sum_{r=1}^{k} \Pr\{V_r\} \frac{B(a_{k+c}-a_{r-c}+k-r-1, k-r-1; q)}{B(m+n-q_{r-c}-r, n-r; q)}. \end{aligned} \tag{6.18}$$

The second equation is obtained on replacing the combination $k + c$ by $k - c$.

Instead of (3.17) we put

$$
\begin{aligned}
u_r &= \Pr\{U_r\} \frac{B(m+n, n; q)}{B(m+n-a_{r+c}-r, n-r; q)} \\
v_r &= \Pr\{V_r\} \frac{B(m+n, n; q)}{B(m+n-a_{r-c}-r, n-r; q)}.
\end{aligned} \tag{6.19}
$$

Then (6 18) becomes

$$
\begin{aligned}
B(a_{k+c} &+ k - 1, k - 1; q) \\
&= \sum_{r=1}^{k} u_r B(a_{k+c} - a_{r+c} + k - r - 1, k - r - 1; q) \\
&\qquad + \sum_{r=1}^{k} v_r B(a_{k+c} - a_{r-c} + k - r - 1, k - r - 1; q).
\end{aligned} \tag{6.20}
$$

This corresponds to the first equation in (3.18). Unfortunately (6.20) is not of the pure convolution type since $a_{k+c} - a_{r+c}$ and $a_{k+c} - a_{r-c}$ are not functions of the two variables $k - r$ and c. The trouble comes from the fact that a_k, as defined by (6.5), is not a linear function of k. It is, however, plausible that we shall commit only an asymptotically negligible error if we omit the brackets in (6 5), that is, if we replace a_k by pk/q. Purely formally (6.20) then reduces to the first equation in (3.18) with

$$
p_k(c) = B\left(\frac{k + cp}{q} - 1, k - 1; q\right). \tag{6.21}
$$

(Here the first argument in the right hand member is no longer necessarily an integer, and the factorials in the definition (2.2) should be interpreted by means of the gamma function.) To the new system (3.18) the considerations of section 3 apply almost word for word: the only difference lies in the new norming (6.14) (which replaces (3.2)) and that instead of (3.26) we shall naturally let $k/n \to t$. Thus the limiting form of Theorem 1 applies to the new system (3.18) with $p_k(c)$ defined by (6.2).

It remains to prove that the formal replacement of (6.20) and the corresponding equation for $-c$, by (3.18) was legitimate. Now all coefficients in (6.20) are of the form $B(\nu, r; q)$, and we have only changed the first argument, ν, adding a variable quantity which in no case exceeds one unit. In passing to the limit we put $k \sim tn$ and $c \sim xn^{\frac{1}{2}}p^{-\frac{1}{2}}$. It follows that we actually use only coefficients $B(\nu, r, q)$ where $\nu \to \infty, r \to \infty$ and $\nu/r \to q$. Accordingly, for $|\vartheta| < 1$ we have $B(\nu + \vartheta, r, q) \sim B(\nu, r; q)$, and it is rather obvious that our system (6.20) is asymptotically equivalent to (3.18).

REFERENCES

[1] A KOLMOGOROFF, "Sulla determinazione empirica di una legge di distribuzione," *Inst Ital Attuari, Giorn.*, Vol. 4 (1933), pp 1–11.

[2] A. KOLMOGOROV, "Über die Grenzwertsatze der Wahrscheinlichkeitsrechnung," *Bulletin [Izvestija] Academie des Sciences URSS*, (1933), pp 363–372

[3] A. KOLMOGOROFF, "Confidence limits for an unknown distribution function," *Annals of Math. Stat*, Vol 12 (1941), pp. 461–463.

[4] R. VON MISES, *Wahrscheinlichkeitsrechnung*, F Deuticke Leipzig und Wien, 1931, p 316 seq

[5] N SMIRNOV, "Ob uklonenijah empiričeskoi krivoi raspredelenija," *Recueil Mathematique (Matematičeskii Sbornik)*, N S Vol 6(48) (1939), pp 3–26

[6] N. SMIRNOV, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin Mathématique de l'Université de Moscou*, Vol 2 (1939), fasc 2.

APPLICATION OF RECURRENT SERIES IN RENEWAL THEORY

By Alfred J. Lotka

Metropolitan Life Insurance Company

Summary. The application of integral equations to renewal theory in population analysis and problems of industrial replacement is beset with certain difficulties which have been particularly discussed by W. Feller (these *Annals* 1941, vol 12 pp. 243–267). Some of these difficulties are avoided if the data of the problem are introduced into the analysis directly in the discontinuous form (tabulated by class intervals) in which they are usually supplied in a concrete case. A numerical example based on population statistics is presented, illustrating how, using discontinuous data, a recurrent series takes the place of the integral equation, and a finite exponential series appears in place of the Heaviside expansion of the previous solution. There is close analogy with the procedure previously presented, but with factorial moments appearing in place of ordinary moments.

The fundamental data being given for values of the replacement function at discrete intervals only, some question arises as to the applicability of the solution as an "interpolation" formula for non-integral values of the time t, and as to the effects of subdividing the class interval of the original data.

In the actual computation of the factorial moments a shift of origin by one-half class interval becomes necessary. An algorithm for effecting this shift is presented.

1. Methodology: Alternatives Available.

All application of mathematics to concrete situations involves a greater or less degree of conventionalisation, a substitution, "in place of intractable reality, of an ideal upon which it is possible to operate."[1]

This conventionalisation may be only such as to do little violence to the concrete data, as for example when, dealing with a large population, we treat the number $N(t)$ of individuals at time t as a continuous variable, knowing perfectly well that strictly speaking it varies by jumps of one unit at a time.[2]

In dealing with any particular concrete case there may be considerable choice as to the mode in which the conventionalisation or idealisation is carried out, and the particular place or step in the scheme at which it is introduced. A good illustration of this is met in the treatment of renewal theory, as applied to human populations or other biological or industrial aggregates.

The majority of authors who have dealt with the subject have set up their fundamental equations in terms of continuous variables. Many have gone fur-

[1] *Nature*, Vol. 110 (1922), p. 764.

[2] If the population is subject to extreme variation in numbers, such that $N(t)$ passes through small values, this disregard of their discontinuity may not be permissible.

ther than this in the process of conventionalisation and have assumed for the renewal function (net reproductivity) some more or less appropriate mathematical expression, such as a Charlier or a Pearson [1] frequency distribution, and have, wherever possible, carried out by standard methods the integrations involved.

Others, while retaining the formulation of the fundamental equations in continuous (infinitesimal) form, have made no specific assumptions regarding the analytical form of the renewal function, and have carried out the numerical integration by one of the established methods available for the approximate integration of arbitrary functions.

But there has also been a minority of authors who deemed it most appropriate, since the data of the problem are actually furnished in tabular (and hence discontinuous) form, to apply from the start discontinuous methods in formulating the fundamental equation for the problem This equation then defines a recurrent series.

The most recent and also the most concise exposition of this approach to the problem is a paper by W. Dobbernack and G. Tietz presented at the Twelfth International Congress of Actuaries, 1940, *Proceedings*, vol. 4, p. 233. These authors, however, do not give any numerical application, and in consequence certain aspects of the analysis are not touched upon by them. A more detailed presentation, including numerical applications, was given by the late S. D. Wicksell[3] who, however, used only roughly approximate data (an over-all average net reproductivity for ages 20 to 44) and also introduced certain linear interpolations which would not be appropriate with more exact data, and which become unnecessary in the numerical operations if moments are introduced as indicated in what follows.

The purpose of the present paper is to exhibit this modification of the method of recurrent series, and at the same time to illustrate its relation to the method which proceeds in terms of a continuous variable, leading to an integral equation.

The $B(t-a)$ women born in the calendar year $(t-a)$, that is, between the times $(t - \frac{1}{2} - a)$ and $(t + \frac{1}{2} - a)$, will be a years old some time during the calendar year t, that is, between $t - \frac{1}{2}$ and $t + \frac{1}{2}$. If their births were evenly distributed over the year $t - a$, so will their birthdays of age a be over the year t, and their average age during that year will be a and the average number of survivors to that age during the year t will be approximately $B(t-a)p(a)$, where $p(a)$ is the probability, at birth, of surviving to age a. If the annual female reproductive rate, counting daughters only, is $m(a)$ at age a, then the $B(t-a)$-$p(a)$ survivors will, during the calendar year t, give birth to $B(t-a)p(a)m(a)$ daughters. If $B(t)$ is the total number of births of daughters in the calendar year t, then evidently, for positive values of t,

$$B(t) = \sum_{1}^{\omega} B(t-a)p(a)m(a), \tag{1}$$

[3] [2]; see also [3].

or, to simplify the notation,

$$B(t) = \sum_{1}^{\omega} c_a B(t - a). \tag{2}$$

Equation (1) or (2) defines a recurrent series of the general form

$$B(t) = c_1 B(t - 1) + c_2 B(t - 2) + \cdots + c_\omega B(t - \omega), \tag{3}$$

where some of the coefficients c may be zero and where ω denotes the upper limit of the reproductive period.

The trial substitution

$$B(t) = Qx^{-t} \tag{4}$$

in (3) gives

$$1 = c_1 x + c_2 x^2 + \cdots + c_\omega x^\omega. \tag{5}$$

The substitution (4) therefore satisfies (3) provided that x is a root of the equation (5) of degree ω for x; and the same is evidently true for the more general substitution

$$B(t) = \sum_{j=1}^{j=\omega} Q_j x_j^{-t}, \tag{6}$$

where x_j, with $j = 1, 2, \cdots \omega$, are the ω roots of (5).

Equation (5) leaves the ω coefficients Q_j indeterminate. In general they appear as arbitrary constants. In any concrete application they may be determined by "initial" conditions; that is, in order to make the problem determinate, it is necessary to be given the values of $B(t)$ for ω successive integral values of t, or some equivalent data.

While, for convenience in description, the analysis has been developed in terms of the year as time unit, the formulae are evidently independent of this choice of unit, provided that the unit employed is adequate for practical application.

Whatever the unit employed, for the direct application of (1) and (3) to a concrete case it is necessary to have the data in such form that values of $p(a)m(a)$ are known for integral values of a. The pertinent statistics do not usually come in that form, the fertility being usually known only for five year age groups, and though it may be sufficient for practical purposes to regard these quinquennial values as representing $p(a)m(a)$ for the midpoint of the group, this yields $p(a) \cdot m(a)$ for fractional values of a, as measured in five year units. We may then proceed as follows: putting

$$x = 1 + y \tag{7}$$

in (5) this becomes

$$
\begin{aligned}
(8) \qquad 1 = &\{c_1 + c_2 + c_3 + \cdots + c_\omega\} \\
&+ \{c_1 + 2c_2 + 3c_3 + \cdots + \omega c_\omega\} y \\
&+ \left\{c_2 + 3c_3 + 6c_4 + \cdots + \frac{\omega(\omega - 1)}{2!} c_\omega\right\} y^2 \\
&+ \left\{c_3 + 4c_4 + 10c_5 + \cdots + \frac{\omega(\omega - 1)(\omega - 2)}{3!} c_\omega\right\} y^3 \\
&+ \cdots \\
&+ \{c_\omega\} y^\omega \\
= &\sum_{h=0}^{h=\omega} \sum_{k=0}^{k=\omega-h} \binom{h + k}{h} c_{h+k} y^h.
\end{aligned}
$$

In application to a particular population, we shall usually have the condition

$$c_a = 0 \quad \text{for} \quad a = 1, 2, \cdots < \alpha$$

where α is the lower limit of the reproductive period.

The expressions in brackets (coefficients of successive powers of y) will be recognized as cumulations S_h of the values of the function c_a, summed backwards to the "diagonal" element c_h, where h is the exponent of y. In terms of moments m of the function c_a, equation (8) can be written

$$(9) \quad 1 = m_0 + m_1 y + \frac{m_2 - m_1}{2!} y^2 + \frac{m_3 - 3m_2 + 2m_1}{3!} y^3 + \cdots + c_\omega y^\omega$$

or, using the symbol $m_{[h]}$ to denote the hth *factorial* moment, equation (9) takes the simple form

$$(10) \qquad 1 = \sum_{h=0}^{h=\omega} \frac{m_{[h]}}{h!} y^h$$

In these expressions the moments m_h and $m_{[h]}$ are those taken about $a = 0$. Actually, the net reproduction rates are given for "semi-values" of a, that is, for values of a which are odd multiples of $\frac{1}{2}$ (using five years as the time unit). By cumulation of these given values moments m_h' and $m_{[h]}'$ about $a = -\frac{1}{2}$ are obtained.[4] From the latter the corresponding functions of the moments about $a = 0$ are obtained by the transformation formulae[5]

$$
(11) \qquad
\begin{aligned}
\frac{m_{[h]}}{h!} &= \sum_{k=0}^{k=h} \frac{(-\frac{1}{2})^{[k]}}{k!} \frac{m_{[h-k]}'}{(h - k)!}, \\
S_h &= \sum_{k=0}^{k=h} \frac{(-\frac{1}{2})^{[k]}}{k!} S_{h-k}'.
\end{aligned}
$$

[4] In these cumulations zero values of c_a for $0 < a < \alpha$ must not be omitted.

[5] In accordance with a customary notation the symbol $(-\frac{1}{2})^{[k]}$ denotes the continued product $-\frac{1}{2}(-\frac{1}{2} - 1)(-\frac{1}{2} - 2) \cdots (-\frac{1}{2} - k + 1)$. In the computation of successive terms, in the sums in the right-hand member of (11), by appropriately laying out the work, advantage is taken of the fact that values of $(-\frac{1}{2})^{[k]}/k!$ for $k = 2, 3 \cdots$ are obtained each from the preceding by multiplying successively by $\frac{3}{4}$, $\frac{5}{6}$, $\frac{7}{8}$, etc, and taking care of the sign, so that fractions with complicated numerators and denominators are avoided.

It will be recalled that in the treatment of the problem of replacement by means of an integral equation,[6] a solution in the form

$$B(t) = \Sigma Q_j' x_j^{-t} = \Sigma Q_j' e^{r_j t}, \tag{12}$$

is obtained, in which the exponential coefficients r_j are the roots of the equation

$$1 = \int_\alpha^\omega e^{-ra} p(a)m(a)\, da = \int_\alpha^\omega x^a p(a)m(a)\, da, \tag{13}$$

i.e.

$$1 = m_0 - m_1 r + \frac{m_2}{2!} r^2 - \frac{m_3}{3!} r^3 + \cdots = \sum_{h=0}^{h=\infty} (-1)^h \frac{m_h}{h!} r^h, \tag{14}$$

in close analogy to equation (10) for y, with the distinction however that in (10) the *factorial* moments take the place of the ordinary moments of (14), and that the series in (10) is finite, terminating at the term in y^ω. There is also an important difference between the characteristic equation (13) and its analogue (5), namely that (5) may admit of negative roots for x, whereas (13) does not admit negative values for x.

2. The constants Q. These are determined by initial conditions, as follows. Equation (2) can be written

$$\begin{aligned} B(t) &= \sum_{a=t}^{a=\omega} c_a B(t-a) + \sum_{a=1}^{a=t-1} c_a B(t-a) \\ &= F(t) + \sum_{a=1}^{a=t-1} c_a B(t-a), \end{aligned} \tag{15}$$

with

$$F(t) = \sum_{a=t}^{a=\omega} c_a B(t-a) \qquad 0 < t \leq \omega \quad \text{and} \quad F(t) = 0 \qquad t > \omega \tag{16}$$

The values of $B(t)$ being given for integral values of t, from $t = -(\omega - 1)$ to $t = 0$, it can be shown that[7]

$$Q_j = \frac{\sum_{t=1}^{t=\omega} F(t) x_j^t}{\sum_{a=1}^{a=\omega} a c_a x_j^a}. \tag{17}$$

---

[6] For a discussion of the limits of applicability of this method See [4].

[7] The reasoning is essentially the same as in the treatment of the problem by integral equations. See [5] and [2, p 39 et seq.].

In the special case that we are tracing the progeny of an initial population all born at the same time, say $B(0)$ births occurring at $t = 0$, so that

$$B(-1) = B(-2) = \cdots = B(-[\omega - 1]) = 0 \tag{18}$$

the expression for Q_j, in view of (5), reduces to a particularly simple form. For if we write the summation in equation (16) in expanded form, we have

$$\begin{aligned}
F(1) &= c_1B(0) + c_2B(-1) + c_3B(-2) + c_4B(-3) + \cdots + c_\omega B(-\overline{\omega - 1}), \\
F(2) &= c_2B(0) + c_3B(-1) + c_4B(-2) + \cdots + c_\omega B(-\overline{\omega - 2}), \\
F(3) &= c_3B(0) + c_4B(-1) + \cdots + c_\omega B(-\overline{\omega - 3}) \\
&\ \ \vdots \\
F(\omega) &= c_\omega B(0).
\end{aligned} \tag{19}$$

If now $B(-1), \cdots, B(-\overline{\omega - 1})$, all vanish, then

$$\sum_1^\omega F(t)x^t = B(0)\{c_1x + c_2x^2 + c_3x^3 + \cdots + c_\omega x^\omega\} \tag{20}$$

$$= B(0) \text{ by (5).} \tag{21}$$

Hence,

$$Q_j = \frac{B(0)}{\sum_{a=1}^{a=\omega} ac_a x_j^a}. \tag{22}$$

In particular

$$B(0) = \sum_{j=1}^{j=\omega} Q_j = B(0) \sum_{j=1}^{j=\omega} \frac{1}{\sum ac_a x_j^a} \tag{23}$$

so that

$$\sum_{j=1}^{j=\omega} \frac{1}{\sum ac_a x_j^a} = 1 \tag{24}$$

The constant $B(0)$ here evidently functions essentially as an arbitrary unit of annual births, and may with this understanding simply be put $= 1$, thereby simplifying the notation. This has been done in what follows, where convenient, especially in the table of constants, Table 3 of the numerical illustration.

The denominator in (17) or (23) can be evaluated for any root x_j of (5) by direct summation if the coefficients c_a are given or have been computed (as indicated below) for integral values of a; or, in a manner similar to that employed in passing from equation (5) to (8), the denominator can be expressed in terms of the cor-

responding roots $y_j = x_j - 1$ of (8) or (10), the cumulations of c_a being replaced by cumulations of ac_a. With the denominator so expressed, the constants Q_j take the form, in obvious analogy to equation (9):

$$Q_j = \frac{\sum_{t=1}^{t=\omega} x_j^t F(t)}{m_1 + m_2 y_j + \frac{m_3 - m_2}{2!} y_j^2 + \frac{m_4 - 3m_3 + 2m_2}{3!} y_j^3 + \cdots} \tag{25}$$

The alternative procedure, to which reference was made in the preceding paragraph, is to operate upon the moments $m_{[h]}$ (taken about the origin O) by a process the inverse of cumulation—which we might term *decumulation*—and in this way to obtain from them the coefficients c_a. The polynomial $\Sigma ac_a x_j^a$ can then be evaluated directly.

The decumulation is readily carried out by an algorithm which suggests itself from the schedule of cumulation. Analytically the relation between the two processes is expressed by the reciprocal sets of transformation formulae:

Cumulation

$$\frac{m_{[h]}}{h!} = \sum_{k=0}^{k=\omega-h} \binom{h+k}{h} c_{h+k} = S_h. \tag{26}$$

Decumulation

$$c_h = \sum_{k=0}^{k=\omega-h} (-1)^k \binom{h+k}{h} \frac{m_{[h+k]}}{(h+k)!}. \tag{27}$$

3. Constants Q associated with complex roots $x = e^{-u+iv}$.

The complex roots x_j give rise to oscillatory terms which, in the special case of the progeny of a cohort of $B(0)$ births, take the form[8]

$$\frac{2B(0)e^{-ut}}{G^2 + H^2} \{G \cos vt - H \sin vt\}, \tag{28}$$

where

$$G = \sum_{a=1}^{a=\omega} ac_a e^{-ua} \cos va \tag{29}$$

and

$$H = \sum_{a=1}^{a=\omega} ac_a e^{-ua} \sin va. \tag{30}$$

These constants may be evaluated directly in this form, or, putting $y = \xi + i\eta$ in the denominator of equation (25), they can be expressed in terms of the roots y_j and the factorial moments obtained by cumulation of ac_a.[9]

[8] The development of these formulae is analogous to that followed in the treatment of the problem by integral equations. See [6]; for the more general case see also [7].

[9] The procedure in this case will be analogous to that followed in the development of equations (90) and (91) in [6].

(a) *Numerical Illustration.* For convenience and to furnish the opportunity for comparison, the same data (United States 1920) were here employed as in the writer's earlier publications in which the problem was treated by the application of an integral equation.

(b) *Cumulation for values of m_h.* The two operations, of (1) cumulating the values of c_a given for semi-values of a, and (2) allowing in the cumulated results

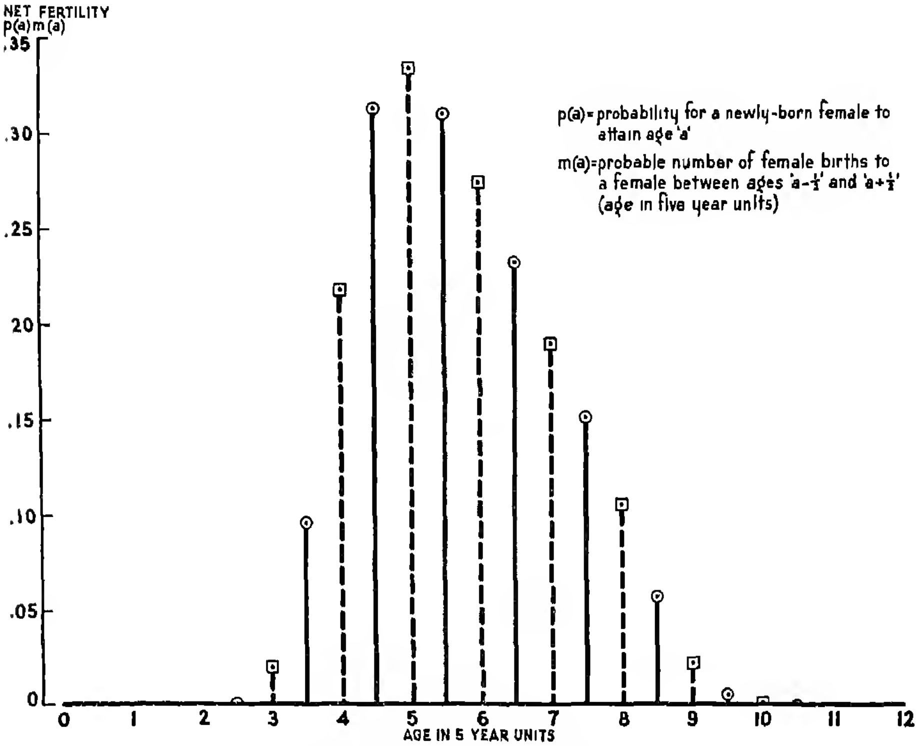


FIG. 1. Net Fertility $p(a)m(a)$ White females, United States, 1920

The verticals drawn in full and centered at mid-ages represent the original data; those drawn in dashed lines and centered at integral ages are interpolated.

for a shift of origin from $a = -\frac{1}{2}$ to $a = 0$, can be conducted in one schedule as in Table 1. Cumulation is first carried out in the usual manner from the bottom line to the diagonal, with the result appearing immediately below the diagonal. From here on the procedure is as in the following example: Starting at the lower right hand corner, we find

$.00780 \times (-\frac{1}{2}) = -.00390$

$.12770 \times (-\frac{1}{2}) = -.06385 \qquad -.06385 \times (-\frac{3}{4}) = .04789$

$.97395 \times (-\frac{1}{2}) = -.48698 \qquad -.48698 \times (-\frac{3}{4}) = .36254$

$.36254 \times (-\frac{5}{6}) = -.30437$

TABLE 1

Computation schedule for values of $\frac{m_{[j]}}{j!} = S_j$ *of net productivity function* $p(a)m(a) = c_a$ *for integral values of age a.**

a in 5-year units	c_a	$m_{[0]}$	$m_{[1]}$	$m_{[2]}/2!$	$m_{[3]}/3!$	$m_{[4]}/4!$	$m_{[5]}/5!$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		1 16635	6 64127	16.04550	24.34106	23.16864	15.05650
			−.58318	.43738	− 36448	.31892	−.28703
0–1	00000	1 16635	7 22445	−3.61223	2.70917	−2.25764	1.97544
1–2	00000	1.16635	6.05810	19.82035	−9 91018	7 43264	−6.19387
2–3	.00040	1 16635	4.89175	13.76225	31.90655	−15.95328	11.96196
3–4	09630	1.16595	3.72540	8.87050	18.14430	33.62800	−16.81400
4–5	.31255	1 06965	2.55945	5.14510	9 27380	15.48370	24.41095
5–6	31025	.75710	1 48980	2.58565	4.12870	6.20990	8.92725
6–7	23170	44685	73270	1.09585	1 54305	2.08120	2.71735
7–8	.15090	.21515	28585	.36315	.44720	53815	.63615
8–9	.05795	06425	.07070	.07730	08405	.09095	09800
9–10	.00615	00630	.00645	00600	.00675	.00690	.00705
10–11	.00015	.00015	00015	.00015	.00015	00015	.00015

a in 5-year units	$m_{[6]}/6!$	$m_{[7]}/7!$	$m_{[8]}/8!$	$m_{[9]}/9!$	$m_{[10]}/10!$	$m_{[11]}/11!$	Factor
(1)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	6 72500	1.99717	.36404	.03483	.00127	.00001	
	.26311	− 24432	.22905	−.21633	.20551	−.19617	−21/22
0–1	−1 77790	1.62974	−1 51333	1.41875	−1.33903	1.27293	−19/20
1–2	5 41964	−4.87768	4.47121	−4.15184	3.89235	−3.67611	−17/18
2–3	−9 97080	8.72445	−7.85201	7.19768	−6.68356	6.26584	−15/16
3–4	12.61050	−10.50875	9.19516	−8.27564	7.58600	−7.04414	−13/14
4–5	−12.20548	9 15411	−7 62843	6.67488	−6 00739	5 50677	−11/12
5–6	12 38595	−6.19298	4.64474	−3.87062	3 38679	−3.04811	−9/10
6–7	3 45870	4 31260	−2 15630	1 61723	−1.34769	1 17923	−7/8
7–8	74135	.85390	97395	−.48698	.36524	−.30437	−5/6
8–9	10520	11255	.12005	.12770	−.06385	.04789	−3/4
9–10	.00720	.00735	00750	.00765	.00780	−.00390	−1/2
10–11	00015	.00015	.00015	.00015	.00015	.00015	

* Figures immediately below the diagonal, obtained by cumulation from the bottom upward of the data in Column 2, are factorial moments about $a = -\frac{1}{2}$. Figures in the top line are factorial moments about $a = 0$ For use of factors in the last column see text.

The several columns are thus completed, and by addition, in each column, of the item immediately below the diagonal, and of all the items above the diagonal, the figures in the top line are obtained. These are the coefficients of equation (10) for y.

(c) *Decumulation.* While it is not necessary to carry out the decumulation, since the entire computation can, if desired, be carried out in terms of y's and m's, there is a considerable interest in noting the values c_a for integral values of a which result from the decumulations of the m's These, together with the original values for semi-values of a, are shown in Table 2 and Fig. 1.

TABLE 2

Values of $c_a = p(a)m(a)$

(1) *for semi-values of a; original data.*

(2) *for integral values of a; computed by cumulation of original data, shift of origin, and decumulation.*

a 5-year units	c_a	a 5-year units	c_a	a 5-year units	c_a
0.0	0	4.0	.21781	8.0	.10607
0.5	0	4.5	.31255	8.5	.05795
1.0	0*	5.0	.33400	9.0	.02268
1.5	0	5.5	.31025	9.5	.00615
2.0	0*	6.0	.27427	10.0	.00116
2.5	00040	6.5	23170	10.5	.00015
3.0	02073*	7.0	.18963	11.0	.00001
3.5	.09630	7.5	.15090		

*The value of c_2 came out negative, namely $-.00570$, and the value of c_1 came out $+.00014$. In the computation of $\Sigma a c_a x^t$ these two values were arbitrarily adjusted to zero, and c_3 was diminished from .02118 to 02073 to make the total $\sum_{t=1}^{11} c_a = 1\,16635$, summing only for integral values of a

4. The roots of equations (5) and (8).

From the prior study already cited, the real positive and three pairs of complex roots for r of the characteristic equation

$$\int_\alpha^\omega x^a\, p(a)m(a)\, da = \int_\alpha^\omega e^{-ra} p(a)m(a)\, da = 1 \tag{31}$$

were known. These were used to indicate the approximate location of the roots of (5) or (8), and more exact values were then obtained by Newton's method of successive approximation. Table 3 shows the values of u, v, etc., corresponding to the new roots

$$\begin{aligned} y &= x - 1 \\ &= e^{-u+iv} - 1 \end{aligned}$$

obtained through equations (8) or (10), and, for comparison the corresponding values obtained in the previous publication from equation (13).[10] The same table also exhibits the remaining roots and values of the constants Q, G, H.

TABLE 3

Constants of the series solution (6) of equation (8), corresponding to the five real and three pairs of complex roots of the characteristic equation (5)

(United States, white females, 1920)

Constants[(1)]	Five Real Roots					Three Pairs of Complex Roots		
				A. Computed on basis of recurrent series				
u	02714*	−1.764†	−3.812†	−17.1†	−94 3†	−.19800	−.44720	−.47587
v	0	0.	0.	0.	0.	1 06498	1.57000	2.40490
G	5.64467	7.73354	−1255.04	(2)	(2)	5 28093	10.45809	7.73103
H	0.	0.	0.	0	0	3.03239	−3 66726	2.00874
$G/(G^2+H^2)$	.17716	12931	−.00080	(2)	(2)	.14241	08515	12117
$H/(G^2+H^2)$	0.	0	0.	0	0	.08177	−.02986	.03148
				B Computed on basis of integral equation‡				
u	.02714					− 1930	−.43655	−.4902
v	0					1.0724	1.5771	2.44245
G	5 64514					5 15351	10.22405	7.40154
H	0.					2.98757	−3.72741	3.45312
$G/(G^2+H^2)$	17715					.14525	.08620	.11095
$H/(G^2+H^2)$	0					.08120	−.03135	.05175

(1) t in five year units
(2) Not computed
* $u_o = \log_e x_o = -\log_e .97322 = .02714$
† Values of x
‡ See [6, p 899]

To determine the remaining four roots, the product of the factors $(y - y_1)$ $(y - y_2) \cdots (y - y_7)$ was divided out of the polynomial of equation (10), rejecting the remainder and leaving a fourth degree equation

$$y^4 + 120\,y^3 + 2590\,y^2 + 14617\,y + 23118 = 0$$

In the subsequent work it turned out that the roots of this were all real, and they were computed by obvious methods. Their values are also shown in Table 3 For the two numerically largest roots great accuracy was not attempted. They introduce terms with very rapid damping and presumably very small values of Q.[11]

[10] The divergence is due in part to details of computation In the earlier publication the curve of fertility $m(a)$ was smoothed by the method of translation, with a Gaussian distribution as basis In the method here presented the raw data were used without smoothing, except such as is inherent in the process of the calculation described.

[11] At any rate, $Q_{10} + Q_{11}$ must be small, since $Q_1 + \ldots + Q_9 = 1.00313$, and according to (24), with the convention that $B(O) = 1$, the sum of all the Q, must be equal to unity.

As a check, in order to be assured that no serious error was introduced in neglecting the remainder after dividing out the factors $(y - y_i)$ up to $(y - y_7)$, the product $\prod_{i=1}^{11} (y - y_i)$ was computed and, after multiplying by a factor to make the absolute terms agree (.16635), was compared with the polynomial of (10). As a further indication, the coefficients of the product Π were "decumulated" to obtain values of coefficients of the corresponding polynomial in x, to

TABLE 4

Coefficients of Powers of y in Equation (10) and in the Product $\prod_1^{11}(y - y_i)$; Also Coefficients of Powers of x in Equation (5)

a	Coefficients of y^a		Coefficients of x^a in Equation (5) Found by Decumulation	
	In Equation (10)	In $\prod_1^{11}(y - y_i)$	Of Column (2)	Of Column (3)
(1)	(2)	(3)	(4)	(5)
0	16635	.16635	−1.00000	−.99915
1	6 64127	6.64072	+ 00014*	+ 00065
2	16 64550	16.64782	−.00057*	−.00432
3	24 34106	24 24197	.02118*	02398
4	23.16840	23 18070	21781	.21774
5	15.05650	15.07338	.33400	.33354
6	6.7250	6.73812	.27427	.27474
7	1 99717	2.00316	.18963	.18882
8	.36404	36555	.10607	.10641
9	.03483	.03501	.02268	.02276
10	.00127	00128	.00116	.00117
11	.00001	00001	00001	.00001

* In computing the denominator of Q according to (22) the values of the coefficients c_1 and c_2 were arbitrarily made zero and the value of c_3 (age 15) was adjusted to 02073 to retain the total $\sum_1^{11} c_j = 1.16635$.

compare with values of c_a The results are shown in Table 4. In view of the fact that the (numerically) highest roots were determined only in first approximation, the agreement is satisfactory.

It is to be noted that instead of applying the solution (6) to compute values of $B(t)$, these latter can, of course, also be obtained directly, by carrying forward step by step the original recurrent series, or, alternatively, the births in successive generations can be computed step by step and the total births obtained by addition. The advantage of the solution (6) is that it enables one, if desired, to obtain $B(t)$ for any value of t without having to compute $B(t)$ for all inter-

vening values of t; also, the solution in an exponential series gives a better idea of the general nature of the process, as well as a direct indication of its asymptotic course for large values of t, when the first term $Q_0x_0^{-t}$ with the positive real root x_0 dominates all others. However this may be, it is interesting to compare

TABLE 5

Synopsis of Results of Computation of $B(t)$ as $\Sigma\ Qx^{-t}$, Column (8), and as $\Sigma B_n(t)$, Column (9), where $B_n(t)$ = Births per Unit of Time in nth Generation at Time t. (Time Unit = 5 years)

	$A = Qx^{-t} = x^{-t}\dfrac{1}{G}$			$A = \dfrac{2e^{ut}}{G^2+H^2}\left\{G\cos vt - H\sin vt\right\}$			
x^* or $u\#$	.97322*	−1.764*	−3.81208*	−.19800#	−.44720#	−.47587#	ΣA
v / t	0	0	0	1.06498	1.57000	2.40490	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	17,716	12,931	−80	28,482	17,030	24,234	100,313
1	18,204	−7,330	21	−415	3,828	−13,781	527
2	18,704	4,156	−6	−19,498	−6,959	3,329	−274
3	19,219	−2,356	1	−15,222	−1,572	2,256	2,326
4	19,748	1,336		1,022	2,844	−3,362	21,588
5	20,291	−757		11,057	646	2,223	33,460
6	20,850	429		8,102	−1,162	−740	27,470
7	21,423	−243		−1,001	−265	−169	19,745
8	22,013	138		−6,248	475	445	16,823
9	22,619	−78		−4,294	109	−344	18,012
10	23,241	44		792	−194	145	24,028
11	23,880	−25		3,519	−45	−1	27,328
12	24,538	14		2,265	79	−55	26,841
13	25,213	−8		−568	18	51	24,706
14	25,907	5		−1,976	−32	−26	23,878
15	26,619	−3		−1,188	−8	4	25,424
16	27,352	1		385	13	6	27,757
17	28,105	−1		1,106	3	−7	29,206
18	28,878	1		620	−5	4	29,498
19	29,673			−251	−1	−1	29,420
20	30,489			−617	2	−1	29,874
21	31,328			−321	1	1	31,008
22	32,191			160	−1	−1	32,349
23	33,076			343			33,419

TABLE 5—*Continued*

t	$B_n(\cdot)$						
	$\Sigma B_n(t)$	*Generations, n*					
		(1)	(2)	(3)	(4)	(5)	(6)
	(9)	(10)	(11)	(12)	(13)	(14)	(15)
0	100,000						
1	0						
2	0						
3	2,072	2,072					
4	21,781	21,781					
5	33,398	33,398					
6	27,472	27,429	43				
7	19,866	18,963	903				
8	16,735	10,607	6,128				
9	17,954	2,268	15,685	1			
10	24,033	116	23,889	28			
11	27,361	1	27,022	338			
12	26,878		24,905	1,973			
13	24,696		18,481	6,214	1		
14	23,851		10,980	12,858	13		
15	25,410		5,345	19,941	124		
16	27,759		2,050	25,030	679		
17	29,219		526	26,316	2,377		
18	29,506		76	23,527	5,897	6	
19	29,414		5	18,092	11,271	46	
20	29,862			12,041	17,579	242	
21	31,000			6,906	23,191	903	
22	32,348			3,381	26,442	2,523	2
23	33,423			1,397	26,426	5,583	17

the result of the computation by means of the exponential series, carried out as set forth above, with the corresponding results of the computation of births in successive generations. This comparison is exhibited in Table 5.

It will be seen that the agreement is good except for the second to fourth items, where perhaps the omission of the terms contributed by the numerically highest roots makes itself felt.

5. Discussion.

(a) *The real roots of the characteristic equation* (5). It can be shown [8] that only one of the real roots for x can be positive, and that the absolute value of any other root must be greater than the positive real root

The negative real roots which make their appearance in the numerical example call for special comment. Practically, the "higher" negative roots are of little importance, at any rate in this example—first because the constants Q with which they are associated are relatively small, second because large absolute values of negative roots imply rapid damping, so that corresponding terms Qx^{-t} very soon become negligible as t increases Thirdly, the determination of these roots would be subject to a wide range of uncertainty, corresponding to the large percentage fluctuations or errors of determination of the values of the functions $p(a)m(a) = c_a$ at the upper end of the reproductive period.

But in theory these negative real roots suggest some pertinent questions. One wonders what would happen to them if the data were given, say, for single years of age, instead of 5-year groups. Instead of an equation of eleventh degree we would then have one of 55th degree. Furthermore, in those cases in which it may be permissible to pass to the limit, so that an integral equation takes the place of (2), negative roots for x would seem to be excluded as they would make the integral in (13) meaningless.

A problem of perhaps little practical importance but of some theoretical interest may arise here, to which reference has also been made by P.H. Leslie in a recent article in *Biometrika*,[12] in connection with a different procedure.

(b) *Effect of finer subdivisions of histogram of* $p(a)m(a)$. The effect of this on equation (5) for x is not obvious at sight, since new coefficients would be inserted *between* previous terms. The effect is more easily understood from a consideration of equation (8) for y. Here finer subdivisions would introduce new terms only beyond the last term originally present. The original terms would not be changed at all *in form*, and those involving only lower moments would be changed but little in *numerical value*, provided that the original histogram were not so coarse as to give inappropriate values even for these lower moments.

The result, then, of finer subdivision of the histogram, would be to change the computed values of the lower roots only in minor degree. But the four negative real roots, depending in considerable measure on the higher terms of (5) or (8), would presumably be materially altered, and might perhaps give place to further complex roots In any case they would be followed by new roots even more remote from practical significance than the original eleven.

(c) *The result as an interpolation formula.* Strictly speaking, the solution (6) of (2) is applicable only for integral values of t. In particular, terms arising out of the negative real roots of (5) for x are obviously not adapted to furnish interpolated values of $B(t)$ for fractional values of t, since fractional powers of

[12] See [9] and [10] For a brief summary and analysis of Leslie's paper [9] see a review signed with the initials WGB in the *Jourl Inst. of Actuaries Student's Soc*, Vol. 4 (1946), Part II The first application of the matrix method to these problems seems to be due to H. Bernardelli, "Population Waves," *Jour. of Burma Research Soc*, Vol. 31 (1941), Part I, p. 1.

negative quantities in general are complex. Over the range of t where the first real root together with the three parts of complex roots adequately describe the process under discussion, these terms alone are, in this sense and to this extent, suitable for interpolation, disregarding the terms corresponding to the other negative roots.

Even less suitable for interpolation purposes, it would seem, would be terms arising from further negative roots that might be introduced by a finer subdivision of the histogram of original data. If we suppose this subdivision carried to great lengths, and if negative roots still appeared under these circumstances, they would give rise to rapidly oscillating positive and negative terms for even and odd integral values of t respectively (the time unit now being a subdivision of the original time unit) with no appropriate interpolation between these integral values

One further point calls for comment. In the process of idealization of the problem discussed, it has been assumed that $p(a)$ and $m(a)$ are independent of time, and the conclusions reached must be construed in the light of this assumption In itself this would hardly call for comment, as it is a matter of common understanding. But the question does arise whether the assumption itself is free from implied internal contradictions.

In a recent publication, P. K. Whelpton[13] has drawn attention to the fact that in times of rapid changes in the birth rate, the assumption of age specific fertilities being held constant at the values observed in a given calendar year may imply that some of the women had more than one *first* child, a logical impossibility.

The data used in the present numerical example are derived from a period of relatively undisturbed birth rate (1920), and do not involve any such conflict. But, in the light of Whelpton's contribution one may ask the broader question whether the computation of an intrinsic rate of natural increase and related parameters based on age specific fertility as observed in one calendar year retain any practical value at all

In answering this question, two considerations will be weighed. First, that *ordinarily* the rates computed in the usual way differ but little from those obtained by taking into account order of birth as in Whelpton's procedure. Secondly, that the computation using over-all values of $m(a)$ for all orders of birth combined is a relatively simple matter based on data commonly available; whereas the more complete treatment of the problem taking into account order of birth is considerably more complicated and often not possible at all for lack of detailed data.

REFERENCES

[1] E C. Rhodes, "Population mathematics," *Roy Stat Soc Jour.*, (1940), pp. 229, 232

[2] S D. Wicksell, "Bidrag till den formella befolkningsteorien," *Statsokonomisk Tidsskrift*, Heft 1–2 (1934), p 1

[13] See [11] Another refinement recently introduced into the measurement of reproductivity is to take into account duration of marriage See Colin Clark and R E. Dyne, *Economic Record* (Australia), June 1946, p. 23

[3] E. J. GUMBEL, "Eine Darstellung statistischer Reihen durch Euler," *Jahresber. der Deutschen Mathematiker Vereinigung*, Vol. 25, p. 257.

[4] W. FELLER, "On the integral equation of renewal theory," *Annals of Math. Stat.*, Vol. 12 (1941), p. 243

[5] A. J. LOTKA, "Contribution to the theory of self-renewing aggregates with special reference to industrial replacement," *Annals of Math. Stat.*, Vol. 10 (1939), p. 11.

[6] A. J. LOTKA, "The progeny of a population element," *Am. Jour. of Hygiene*, Vol. 8 (1928), p 892.

[7] A. J. LOTKA, "The progeny of an entire population," *Annals of Math. Stat.*, Vol. 13 (1942), pp 117–118.

[8] W. DOBBERNACK AND G. TIETZ, *Twelfth International Congress of Actuaries*, Vol 4 (1940), p. 233

[9] P. H LESLIE, "On the use of matrices in certain population mathematics," *Biometrika*, Vol. 33 (1945), pp 202 and 206.

[10] E. G. LEWIS, "On the generation and growth of a population," *Sankhya*, Vol. 6 (1942), Part I, p. 93 et seq.

[11] P. K. WHELPTON, "Reproduction rates adjusted for age, parity, fecundity, and marriage, *Am. Stat. Assn Jour.*, Vol. 41 (1946), p. 501.

SOLUTION OF EQUATIONS BY INTERPOLATION

W. M. KINCAID

University of Michigan

Introduction and summary. The present paper deals with the numerical solution of equations by the combined use of Newton's method and inverse interpolation. In Part I the case of one equation in one unknown is discussed. The methods described here were developed by Aitken [1] and Neville [2], but do not seem as widely known as they should be, perhaps because the original papers are not readily available. (A short summary of Aitken's work will be found in a recent paper by Womersley [3].) Mention should also be made of an interesting paper by Spoerl [4], which treats the same problem from a somewhat different viewpoint.

In Part II these methods are extended to sets of simultaneous equations

PART I. EQUATIONS IN ONE UNKNOWN

1. Nature of the problem. We first consider the problem of locating, to any desired degree of accuracy, a real root x_0 of an equation of the form

$$y(x) = 0 \tag{1}$$

where $y(x)$ is assumed to be analytic in an interval containing the root in question. Since we shall not be concerned here with the necessary preliminary work of separating the roots, etc., we may suppose that x_0 is known to lie within a given interval that contains no zeros of $y'(x)$. (Multiple roots are thus excluded; but of course any such root is a simple root of an equation obtained from (1) by differentiation, and the methods described below can be applied to this equation.)

2. Aitken's method of interpolation. The method to be described, which may be regarded as a generalization of Newton's, depends on the use of inverse interpolation. It is therefore desirable to recall a few points from the theory of interpolation before proceeding further.

Let f be a function such that $f(t)$ is known for $t = t_1, t_2, \cdots, t_n$. Then the Lagrange interpolating polynomial $f_{12\cdots n}(t)$ is defined by

$$\begin{aligned} f_{12\cdots n}(t) &= f(t_1)\frac{(t-t_2)(t-t_3)\cdots(t-t_n)}{(t_1-t_2)(t_1-t_3)\cdots(t_1-t_n)} \\ &\quad + f(t_2)\frac{(t-t_1)(t-t_3)\cdots(t-t_n)}{(t_2-t_1)(t_2-t_3)\cdots(t_2-t_n)} \\ &\quad + \cdots + f(t_n)\frac{(t-t_1)(t-t_2)\cdots(t-t_{n-1})}{(t_n-t_1)(t_n-t_2)\cdots(t_n-t_{n-1})}. \end{aligned} \tag{2}$$

We note that

$$f_{12}(t) = f(t_1)\,\frac{t - t_2}{t_1 - t_2} + f(t_2)\,\frac{t - t_1}{t_2 - t_1} = \frac{\begin{vmatrix} f(t_1) & t - t_1 \\ f(t_2) & t - t_2 \end{vmatrix}}{t_1 - t_2},$$

(3)

$$f_{123}(t) = \frac{\begin{vmatrix} f_{12}(t) & t - t_1 \\ f_{23}(t) & t - t_3 \end{vmatrix}}{t_1 - t_3}, \cdots, f_{123\cdots n}(t) = \frac{\begin{vmatrix} f_{123\cdots n-1}(t) & t - t_1 \\ f_{23\cdots n}(t) & t - t_n \end{vmatrix}}{t_1 - t_n},$$

so that $f_{123\cdots n}(t)$ can be evaluated for any given value t_0 of t by a succession of linear interpolations. It is convenient to arrange the work in a table like the following ($n = 4$):

TABLE Ia

t	$f(t)$	I	II	III	$Parts$
t_1	$f(t_1)$				t_0-t_1
		$f_{12}(t_0)$			
t_2	$f(t_2)$		$f_{123}(t_0)$		t_0-t_2
		$f_{23}(t_0)$		$f_{1234}(t_0)$	
t_3	$f(t_3)$		$f_{234}(t_0)$		t_0-t_3
		$f_{34}(t_0)$			
t_4	$f(t_4)$				t_0-t_4

This form is well adapted for machine computation, for each denominator $t_i - t_j = (t_0 - t_j) - (t_0 - t_i)$ automatically appears in one set of counters when the corresponding numerator is obtained in the other

If $f'(t)$ is known at one or more of the given points, this information is readily fitted into the scheme. For we see that

(4) $$f_{11}(t) \equiv \lim_{t_2 \to t_1} f_{12}(t) = f(t_1) + (t - t_1)f'(t_1)$$

and all that is necessary is to repeat certain entries in Table Ia and to fill in column I by using (4) as indicated in Table Ib. The extension to higher derivatives is obvious

TABLE Ib

t	$f(t)$	I	II	III	$Parts$
t_1	$f(t_1)$				t_0-t_1
		$f_{11}(t_0)$			
	$f(t_1)$		$f_{112}(t_0)$		t_0-t_1
		$f_{12}(t_0)$		$f_{1123}(t_0)$	
t_2	$f(t_2)$		$f_{123}(t_0)$		t_0-t_2
		$f_{23}(t_0)$		$f_{1233}(t_0)$	
t_3	$f(t_3)$		$f_{233}(t_0)$		t_0-t_3
		$f_{33}(t_0)$			
	$f(t_3)$				t_0-t_3

In applying the above to obtaining the root x_0 of (1), we must suppose that $y(x)$ is tabulated or can be computed for a set of values of x in the neighborhood of x_0. What we do not know is the value of x corresponding to $y = 0$. It is therefore convenient to regard x as a function of y whose value is known at certain points and then interpolate to get $x_0 = x(0)$. That is, we let y take the place of t and x that of $f(t)$ in the preceding discussion, while 0 replaces t_0. The work is slightly simplified by the fact that the column of "parts" becomes identical with the left-hand column which contains the y's and can therefore be omitted.

3. Application to an example. The procedure will be most clearly indicated by an example. Consider the equation

$$y = x^4 + 2x^3 - 5x^2 - 8x + 1 = 0, \tag{5}$$

which has a root between 0 and 1. (If the root were located elsewhere, it would be desirable to shift it to this interval in order to simplify the computation of y.)

The work of evaluating this root to ten places is summarized in Table II, and explained below. In the first column, the numbers in parentheses are values of $\frac{dy}{dx}$, and the other numbers are values of y, corresponding to the values of x in the second column.

TABLE II

y	x	I	II	III
1.000 000 000 000	0 000			
(−8 000 000 000)	0.000	0 125 000 000 00		
0.152 100 000 000	0 100	0 117 938 436 13	0 116 671 702 00	
−0.001 054 385 279	0 117	6 882 964 17	884 075 87	
(−9 081 459 548)	0 117	3 896 94	3 890 52	0 116 883 877 01
0 008 022 855 936	0 116	3 842 98	90 67	90 68
(−9 073 020 416)	0.116	4 254 15	90 74	90 68

$x_0 = 0\ 116\ 883\ 890\ 7$

The procedure is as follows. Taking $x = 0$ as a first approximation to x_0, we find that $y(0) = 1$, $y'(0) = -8$, and record this data in the y and x columns of the table. Note that for convenience, the value of $y'(0)$ takes the place of a blank entry in Table Ib. We now apply (4), which here takes the form

$$x_{11}(0) = x\Big|_{y=1} + (0 - 1)\frac{dx}{dy}\Big|_{y=1} = 0 + \frac{-1}{\left.\frac{dy}{dx}\right|_{x=0}} = 0 + \frac{-1}{-8} = 0.125 \tag{6}$$

and enter the result in column I. Note that this is equivalent to one step of Newton's method.

In view of (6) we take $x = 0.1$ for our next approximation and apply (3) to obtain the second entry in column I and the first in column II. This last suggests $x = 0.117$ for our next trial value. (We do not compute $y'(0.1)$, as little

would be gained by doing so, and the time is better spent in going ahead as indicated.) Finding $y(0.117)$ and filling in the table gives us the root to six places.

4. Employment of tables. Continuing in the same line, it would seem natural to take $x = 0.116884$ at the next step, and doing so would lead to the most rapid convergence. But another consideration enters. Up to this point the values of y were computed with the aid of the WPA Table of Powers, which is limited to three places in the argument. Rather than going to the extra labor of evaluating $y(.116884)$, we proceed as indicated in the table, using $y'(.117)$, $y(.116)$ and $y'(.116)$, and stopping when the values of x in the last column agree to the desired number of places

This point has been dwelt on because it is likely to arise whenever tables are used in evaluating $y(x)$. In the example just given, to be sure, we had a certain freedom of choice; but if $y(x)$ is not algebraic, direct computation may be quite impractical. It may be noted that in such cases the method of inverse interpolation is not only faster than the simple Newton's method but is capable of giving more accurate results.

The error in the final result can be estimated from the standard formula for the error of interpolation, but this may be awkward because it requires the evaluation of higher derivatives of x with respect to y. In practice it is generally safe to rely on agreement of different interpolated values, and of course the result may be checked by substitution in the original equation. One simple point is worth noting, however—if the error in the original column of x's is $O(\epsilon)$, that in the successive columns to the right is $O(\epsilon^2)$, $O(\epsilon^3)$, etc.

5. Applicability of the method. Although the example we have presented is algebraic, the method is, of course, equally applicable to transcendental equations. Moreover, it can be used, theoretically at least, to yield complex as well as real roots. The sole difficulty is that the numerical work becomes cumbersome in this case, how serious it is depends on the type of computing machines used. If the equation is algebraic, Bernoulli's [5], [6] and Graeffe's [7] methods are applicable. In fact, they are likely to be the most effective since they do not require prior knowledge of a first approximation to the root. If the alternative procedure of replacing the equation by two simultaneous equations for the real and imaginary parts of the root is decided upon, the methods described in the next section may prove useful.

PART II. SETS OF SIMULTANEOUS EQUATIONS

6. Two equations; general considerations. It is natural to take up next the problem of finding the simultaneous solutions of two equations in two unknowns. Let these equations be

$$u(x, y) = 0, \qquad v(x, y) = 0, \tag{7}$$

where u and v are analytic functions of x and y.

If we had a general method of interpolation of functions of two independent variables, the problem could be solved in a fashion similar to that used in the preceding section. That is, u and v would be computed for values of x and y near the desired ones; then x and y would be regarded as functions of u and v and interpolations would be performed to obtain the values corresponding to $u = v = 0$.

It is easy to set up interpolating functions in a variety of ways, but the author has found none that are satisfactory for the problem in hand. Note that what is required is to determine the value of a function at any point in the plane, given its values at a set of fixed points The most obvious idea is to use polynomials of the least possible degree for this purpose, as is done in the case of a single variable. In this case, however, the coefficients of a polynomial of the nth degree are determined by its values not at $n + 1$ but at $\frac{(n+1)(n+2)}{2}$ points; thus if a function is given at 5 points, no unique quadratic interpolating polynomial can be constructed. What is worse, even if a function is given at 6 points, say, the quadratic polynomial determined will in general have large coefficients and take on unreasonable values if all the points happen to lie close to a common conic. Other schemes considered by the author have similar drawbacks, though the possibility of course remains of finding a suitable one by further research.

The problem can also be handled, at least in principle, by eliminating one of the variables; but, apart from the difficulty of carrying this out in practice, the resulting single equation is likely to be more complicated in form than the original two. If so, solving it may require more computation than would be involved in attacking the original equations directly by the methods described below. So far is this true that even when a single equation is given in the first place it may be advantageous to replace it by a set of simpler equations.

7. Newton's Method. Although a direct extension of the method of inverse interpolation is not presently available, Newton's method may be suitably generalized for this case.

Starting with equations (7), we set up the auxiliary variables

$$X = uv_y - vu_y, \qquad Y = uv_x - vu_x, \tag{8}$$

where the subscripts denote partial derivatives; $u_x = \dfrac{\partial u}{\partial x}$, $u_y = \dfrac{\partial u}{\partial y}$, etc.

We have

$$\begin{aligned} \frac{\partial X}{\partial x} &= u_x v_y - v_x u_y + uv_{xy} - vu_{xy}, \qquad & \frac{\partial X}{\partial y} &= uv_{yy} - vu_{yy}, \\ \frac{\partial Y}{\partial x} &= uv_{xx} - vu_{xx}, & \frac{\partial Y}{\partial y} &= u_y v_x - v_y u_x + uv_{xy} - vu_{xy}. \end{aligned} \tag{9}$$

For $u = 0, v = 0$, equations (8), (9) reduce to

$$X = Y = \frac{\partial X}{\partial y} = \frac{\partial Y}{\partial x} = 0, \qquad \frac{\partial X}{\partial x} = -\frac{\partial Y}{\partial y} = J, \tag{10}$$

where J is the Jacobian of u and v with respect to x and y.

Equations (10) will hold approximately for values of x and y near those satisfying equations (7). That is, in the neighborhood of a solution X can be regarded as a function of x alone and Y as a function of y alone. Then if $x = x_0$, $y = y_0$ is the desired solution, (x_1, y_1) is a point in its neighborhood, and $x_1 = X(x_1, y_1)$, $Y_1 = Y(x_1, y_1)$, $J_1 = J(x_1, y_1)$, we have

$$x_0 \sim x_1 - \frac{X_1}{J_1}, \qquad y_0 \sim y_1 + \frac{Y_1}{J_1}. \tag{11}$$

Also if (x_2, y_2) is another point near (x_0, y_0),

$$x_0 \sim \frac{x_1 X_2 - x_2 X_1}{X_2 - X_1}, \qquad y_0 \sim \frac{y_1 Y_2 - y_2 Y_1}{Y_2 - Y_1}. \tag{12}$$

Relations (11) and (12) can be used to obtain successive approximations to the solution. Use of these relations corresponds to employing Newton's method and linear interpolation for the solution of one equation in one unknown.

As a first example we consider the equations

$$\begin{aligned} u &\equiv x^2 + xy + y^2 - 3 = 0 \\ v &\equiv x^2 y + y^2 - 1 = 0. \end{aligned} \tag{13}$$

We have

$$\begin{aligned} u_x &= 2x + y \qquad & u_y &= x + 2y \\ v_x &= 2xy \qquad & v_y &= x^2 + 2y. \end{aligned} \tag{14}$$

Drawing a rough graph indicates a solution near (1, 1). We evaluate u, v, etc., at this point as shown in Table III. Using (11) we get (2, 0) for a second approximation, and proceed as before. We can now use both (12) and (11) to get new approximations; they are (1.33, 0.57) and (1.25, 0.50), and are entered in the last two columns of the table We therefore try (1.3, 0.5) next, and continue in this fashion until the desired accuracy is attained. Both (11) and (12) are used at each step and the values obtained are entered in the last two columns. The entries in the numbered rows are obtained by using (11), the others by using (12). The number of places to take in each succeeding step is judged from the agreement shown.

Table IV indicates the process of finding a second solution of (11) by the same method. The convergence is very rapid in this case, mainly because the first guess is fairly close.

	x	y	u	v	u_x	v_x	u_y	v_y
1	1.0	1.0	0.0	1.0	3 0	2.0	3.0	3.0
2	2 0	0.0	1 0	−1 0	4 0	0 0	2.0	4 0
3	1 3	0 5	− 41	095	3 1	1 3	2.3	2.69
4	1.5	0 4	.01	06	3 4	1.2	2.3	3.05
5	1 51	0.37	− 0243	− 019463	3 39	1 1174	2 25	3.0201
6	1 5138	0 3750	$-.0_3 10956$	$-0_4 28585$	3.4026	1.13535	2 2638	3.04159044
7	1 5138345	0 37499651	$-.0_7 698799749$	$-.0_7 3062799647$	3.40266551	1.1353653084	2 26382752	3 0416879134

	J	X	Y	(x)	(y)
1	3.0	−3.0	−3 0	2.0	0.0
				1 333	0 5714
2	8.0	6 0	4 0	1.250	0 500
				1.4263	.4143
3	5 349	−1 3214	−.8275	1.54704	34530
				1.51771	.36979
4	7 61	−.1075	−.192	1.51413	.37477
				1 513799	.375046
5	7 723989	−.02959668	.03882675	1.51383179	.37503677
				1 51383479	.374996509
6	7.779110301	$-.0_3 2685259256$	$-.0_4 27125625$	1.51383451188	.374996513
				1 51383451841017	.374996513197826
7	7 779575324	$-0_6 1432165737$	$0_7 2487752800$	1.5138345184093048	.3749965131978003

x_0 = 1 51383451 84093
y_0 = 0.3749965131978

TABLE IV

	x	y	u	v	u_x	v_x	u_y	v_y
1	−1.0	−1.0	0.0	−1.0	−3.0	2.0	−3.0	−1.0
2	−.7	−1.3	.09	.053	−2.7	1.82	−3.3	−2.11
3	−.699	−1.274	$0_2 2203$	$.0_3 598326$	−2.672	1.781052	−3.247	−2.059399
4	−.69877	−1.27351	$-.0_5 21843$	$-.0_5 1122373279$	−2.67105	1.7797811654	−3.24579	2.0587404871

	f	X	X	(x)	(y)
1	9.0	−3.0	−3.0	−.667	−1.33
				−.6984	−1.2721
2	11.703	−.015	.3069	−.69872	−1.27378
				−.69879	−1.273524
3	11.28578997	$-.0_2 2594091475$	$.0_5 5522384628$	−.698770145	−1.27351068
				−.698770075686	−1.27351061019
4	11.27579469	$-0_6 8539188805$	$-.0_5 6885491146$	−.69877007573026	−1.27351061064354

$x_0 = -0.69877007573$
$y_0 = -1.27351061064$

8. Inverse interpolation. In the preceding section, attention was drawn to the difficulty that may arise when tables, necessarily limited to a certain number of places in the argument, are used in the computation. In the example just discussed the values of u and v were easily computed directly to the number of places wanted. But a glance at the work will show that if we had been limited in computing u and v to values of x and y having, say, two decimal places, the solutions could have been carried to four places only.

The device adopted in the preceding section was to use quadratic and cubic interpolates to secure greater accuracy, and it might occur to us to try the same idea here. But for such an interpolation to be strictly valid, equations (10) would have to hold identically. Since they hold only approximately, an error is introduced which, in general, is of the same order of magnitude as the error in linear interpolation. Thus continuing the interpolation would not improve the results

However, this very situation suggests a way out. For suppose we give x a constant value x_1, and compute X and Y for a number of values of y. For $x = x_1$, both X and Y can be regarded as functions of y alone, or we can regard X and y as functions of Y. Doing so, we can interpolate to any number of stages to find values of X and y corresponding to $Y = 0$; call these X_1, y_1. Assigning x other constant values $x_2, x_3, \cdots, x_m$, we repeat the process, getting a set of values $X_2, \cdots X_m$ and $y_2, \cdots, y_m$, all corresponding to $Y = 0$. Now along the curve $Y = 0$ we can regard x and y as functions of X; performing one more interpolation, we obtain the desired values of x, y corresponding to $X = Y = 0$. The error in the final result can be estimated from the errors in the interpolations, and is of the same order of magnitude as the greatest of these.

It will be noted that we did not refer to the definitions of X and Y in describing this procedure. Any pair of independent (analytic) functions X' and Y' having the property that $X' = Y' = 0$ when $u = v = 0$ could be used. However, it is convenient to choose them so that $\dfrac{\partial X'}{\partial y}$ and $\dfrac{\partial Y'}{\partial x}$ are small Probably the simplest course is to set

$$X' = a_1u + b_1v, \qquad Y' = a_2u + b_2v,$$

where a_1, a_2, b_1, b_2 are constants such that

$$\frac{a_1}{b_1} \sim -\frac{v_y}{u_y}, \qquad \frac{a_2}{b_2} \sim -\frac{v_x}{u_x}$$

Let us apply this procedure to the example we have already worked (Table III). Suppose we wish to use values of x and y having not more than two decimal places. Within this restriction, we can still carry through the first few steps indicated in Table III to ascertain that $x_0 \sim 1.514$, $y_0 \sim 0.375$ where (x_0, y_0) is the desired solution. At the point (1.51, 0.37) we have

$$X = 3.0201u - 2.25v, \quad Y = 1.1174u - 3.39v$$

TABLE V

x	y	u	v	X'	Y'	y	y_I	y_{II}	y_{III}
1 50	0.36	−.0804	− 0604	−.1404	1008	0.36			
	0.37	−.0581	− 0306	−.1406	.0337	0 37	.3750223546.9		
	0.38	−.0356	−.0006	− 1406	− 0338	0.38	49925925 9	3759999662 8	
	0.39	−.0129	.0296	− 1404	−.1017	0.39	50220913 1	49999345 9	.3750000007.3
1.51	0.36	−.0467	−.049564	−.038108	.101992	0.36			
	0.37	−.0243	− 019463	−.038811	.034089	0.37	.3750202494 7		
	0.38	−.0017	.010838	−.039314	−.034124	0 38	49908495 9	.3749982346 4	
	0.39	−.0211	.041339	−.039617	−.102917	0.39	50200136.8	81060.2	.3749981706 2
1 52	0.36	−.0128	−.038656	.064768	.103168	0.36			
	0.37	.0097	−.008252	063566	.034456	0.37	.3750145534 9		
	0.38	.0324	.022352	062544	−.034656	0.38	49855307 3	.3749928285.4	
	0.39	.0553	.053156	.061732	−.104168	0.39	50143860.0	27029.2	.3749927660.3
1 53	0.36	.0213	−.027676	.168228	.104328	0.36			
	0.37	.0439	.003033	166501	.034801	0.37	.3750053935.8		
	0.38	.0667	.033942	.164974	−.035126	0.38	49767614.7	.3749839733 9	
	0.39	.0897	.065051	.163647	−.105453	0.39	50053322.3	38506.9	.3749839123.7

Y'	X'	X'_I	X'_{II}	X'_{III}
.1008	−.1404			
.0337	−.1406	−.1407004470.9		
−.0338	−.1406	6000000 0	−.1406252237.1	
−.1017	−.1404	6995581.7	47792 5	− 1406250024.7
101992	−.038108			
.034089	−.038811	−.03916392353 7		
− 034124	−.039314	06203973.4	−.03908763223.7	
−.102917	− 039617	16310641.4	718653 1	−.03908741039 0
.103168	.064768			
.034456	063566	.06294823611.5		
−.034656	.062544	305146428.9	06302550744 8	
−.104168	.061732	294883185 6	595425 3	.06302572977.3
.104328	.168228			
.034801	166501	1656365685 3		
−.035126	.164974	7410485.2	1657147318.6	
−.105453	.163647	6367924 1	51796 2	.1657149545 4

X'_{III}	y_{III}	$y_{I'}$	$y_{II'}$	$y_{III'}$
− 1406250024 7	.3750000007.3			
−.03908741039.0	.3749981706.2	.3749974661.1		
06302572977.3	.3749927660 3	61018.2	3749965240 4	
.1657149545.4	.3749839123.7	81999 8	65022 6	3749965140 4

X'_{III}	x_{III}	$x_{I'}$	$x_{II'}$	$x_{III'}$
−.1406250024.7	1.50			
−.03908741039.0	1 51	1.5138495506 5		
.06302572977.3	1.52	278531 3	1.5138345680.7	
.1657149545.4	1.53	624787.5	44615.8	1 5138345191.9

Noting the ratios of the coefficients of u and v, we select

$$X' = 4u - 3v, \quad Y' = u - 3v.$$

Next we evaluate X' and Y' for the 16 points having x-coordinates 1.50, 1.51, 1.52, 1.53 and y-coordinates 0.36, 0.37, 0.38, 0.39, as shown in Table V. Starting with the four points for which $x = 1.50$, we interpolate to find the values of y and X' corresponding to $Y' = 0$; they are $y_1 = .3750000007$, $X_1' = -.1406250025$. We proceed in the same way with the points corresponding to the other values of x; the results, as shown, are $y_2 = .3749981706$, $X_2' = -.03908741039$; $y_3 = .3749927660$, $X_3' = .06302572977$; $y_4 = .3749839124$, $X_4' = .1657149545$. (The extra digits given in Table V are to take care of rounding-off.) Finally, using these values, we interpolate to find the values of x and y corresponding to $X' = 0$, and get

$$x = 1.5138345192, \quad y = .3749965140$$

Comparing these results with those obtained earlier, we see that they are in error by about 1 unit in the ninth place; a distinct improvement over the four correct places that could have been secured without using this device. Note that if we had not had our earlier results for comparison, a check could have been obtained by carrying through the interpolation in the reverse order; i.e., starting with fixed values of y and finding values of x and Y' corresponding to $X' = 0$.

As in the case of one equation in one unknown, derivatives could be brought into the interpolation scheme, permitting greater accuracy with fewer points. But the derivatives needed would be $\frac{\partial x}{\partial X'}, \frac{\partial x}{\partial Y'}, \frac{\partial^2 x}{\partial X' \partial Y'}$, etc., and the general setup would be rather awkward, so that extra labor would probably be required.

9. Three or more equations. The methods discussed in this section are readily extended to the solution of three or more simultaneous equations in an equal number of unknowns. For example, if we are given three equations of the form

$$u(x, y, z) = 0, \quad v(x, y, z) = 0, \quad w(x, y, z) = 0,$$

we define new variables

$$X = \begin{vmatrix} u & v & w \\ u_y & v_y & w_y \\ u_z & v_z & w_z \end{vmatrix}, \qquad Y = \begin{vmatrix} u_x & v_x & w_x \\ u & v & w \\ u_z & v_z & w_z \end{vmatrix}, \qquad Z = \begin{vmatrix} u_x & v_x & w_x \\ u_y & v_y & w_y \\ u & v & w \end{vmatrix}$$

which are analogous to the X and Y of (8); from this point on the work is practically the same as before.

REFERENCES

[1] A. C. Aitken, "On interpolation by iteration of proportional parts, without the use of differences," *Edinb Math. Soc Proc.*, ser 2, Vol. 3 (1932), pp. 56–76.

[2] E. H Neville, "Iterative interpolation," *Indian Math. Soc. Jour*, Vol 20 (1934), pp 87–120.

[3] J R Womersley, "Scientific computing in Great Britain," *Mathematical Tables and aids to Computation*, Vol. 2 (1946), pp 110–117.

[4] C. A. Spoerl, "Solving equations in the machine age," *Amer Inst. Actuar. Record*, Vol. 31, Part I (1942), pp. 129–149

[5] T. C. Fry, "Some numerical methods for locating roots of polynomials," *Quart. Appl. Math.*, Vol 3 (1945), pp. 89–105.

[6] W. M Kincaid, "Numerical methods for finding roots and vectors of matrices," *Quart. Appl. Math.*, Vol 5 (1947), pp 320–345

[7] E Bodewig, "On Graeffe's method for solving algebraic equations," *Quart. Appl. Math.* Vol. 4 (1946), pp 177–190.

ESTIMATION OF A PARAMETER WHEN THE NUMBER OF UNKNOWN PARAMETERS INCREASES INDEFINITELY WITH THE NUMBER OF OBSERVATIONS

By Abraham Wald

Columbia University

Summary. Necessary and sufficient conditions are given for the existence of a uniformly consistent estimate of an unknown parameter θ when the successive observations are not necessarily independent and the number of unknown parameters involved in the joint distribution of the observations increases indefinitely with the number of observations. In analogy with R. A. Fisher's information function, the amount of information contained in the first n observations regarding θ is defined. A sufficient condition for the non-existence of a uniformly consistent estimate of θ is given in section 3 in terms of the information function. Section 4 gives a simplified expression for the amount of information when the successive observations are independent.

2. Introduction. J. Neyman has recently treated the following estimation problem[1]: Let $X_1, X_2, \cdots$, etc. be a sequence of independent chance variables the distribution of each of which depends on some unknown parameters. Two kinds of parameters are distinguished, structural and incidental parameters. A parameter θ is called structural if there exists an infinite subsequence of the sequence $\{X_i\}$ such that the distribution of each of the chance variables in the subsequence depends on θ. Any parameter which is not structural is called incidental. Neyman has considered the case when there are a finite number of structural parameters, say $\theta_1, \cdots, \theta_s$ and an infinite sequence $\{\xi_i\}$, ($i = 1, 2, \cdots$, ad inf.), of incidental parameters. He has studied the problem of consistent and efficient estimation of the structural parameters and has obtained several interesting results. He has shown, among others, that the maximum likelihood estimate of a structural parameter θ need not be consistent, even when consistent estimates of θ exist. Neyman has also given a method for obtaining consistent estimates of the structural parameters. This method, however, is applicable only under certain restrictive conditions.

In this paper we shall consider a more general case than that treated by Neyman, but we shall concentrate on one aspect of the problem, namely that of the existence of consistent estimates.

Let $\{X_i\}$, ($i = 1, 2, \cdots$, ad inf.), be a sequence of chance variables, not necessarily independent of each other. It is assumed that for each n the chance variables $X_1, \cdots, X_n$ admit a joint probability density function $p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)$ where $\theta, \xi_1, \xi_2, \cdots$, etc. are unknown parameters.[2]

[1] Address given by J. Neyman at the meeting of the Institute of Mathematical Statistics in Atlantic City, January, 1947.

[2] While θ is assumed to be a real variable, we admit ξ_i to be a finite dimensional vector, i.e., $\xi_i = (\xi_{i1}, \cdots, \xi_{ik_i})$ where k_i may be any finite positive integer.

We shall require that the consistency relations among the density functions $p_1, p_2, \cdots$, etc. be fulfilled, i.e.,

$$\int_{-\infty}^{+\infty} p_{n+1}\, dx_{n+1} = p_n, \qquad (n = 1, 2, \cdots, \text{ad inf.}). \tag{1.1}$$

It should be remarked that it is not postulated that p_n actually depends on all the parameters that appear as arguments in p_n. It is merely assumed that p_n does not depend on any parameter that does not appear as an argument in p_n, i.e., p_n does not depend on ξ_i for any $i > n$. It follows, however, from (1.1) that if p_n depends on a parameter ξ, then also p_m depends on ξ for any $m > n$.

Neyman's definition of structural and incidental parameters can be extended to the case of dependent observations considered here by saying that the distribution of X_i does not depend on a parameter ξ if and only if the conditional distribution of X_i for any given values of $X_1, \cdots, X_{i-1}$ does not depend on ξ. It is not postulated that each of the parameters $\xi_1, \xi_2, \cdots$, etc. is incidental; some of them may be structural. We shall not make an explicit distinction between structural and incidental parameters, since for the purposes of the present paper this does not seem to be necessary.

In this paper we shall deal with the problem of formulating conditions under which a uniformly consistent estimate of θ exists. A statistic $t_n(x_1, \cdots, x_n)$ is said to be a uniformly consistent estimate of θ if for any positive δ

$$\lim_{n=\infty} \text{prob.} \{|t_n - \theta| < \delta\} = 1 \tag{1.2}$$

uniformly in θ and the ξ's.

In section 2 a necessary and sufficient condition is given for the existence of a uniformly consistent estimate of θ. In section 3 the amount of information supplied by the first n observations concerning θ is defined. It is then shown that if the amount of information is a bounded function of n over a non-degenerate θ-interval, no uniformly consistent estimate of θ exists. Section 4 gives a simplified formula for the amount of information in the case when the X's are independently distributed.

2. A necessary and sufficient condition for the existence of a uniformly consistent estimate of θ. In deriving a necessary and sufficient condition for the existence of a uniformly consistent estimate of θ, use will be made of some results contained in a publication of the author [1] dealing with statistical decision functions which minimize the maximum risk. In [1] it is assumed that the domain of each of the unknown parameters is a closed and bounded set and that p_n is continuous jointly in all of its arguments. Thus, in order to be able to use the results obtained in [1], we shall have to make the same assumptions here. In what follows we shall, therefore, assume that each of the parameters $\theta, \xi_1, \xi_2, \cdots$, etc is restricted to a finite closed interval and that p_n is a continuous function of $x_1, \cdots, x_n, \theta, \xi_1, \cdots, \xi_n$.

Let $[a, b]$ $(a < b)$ be the θ-interval to which the values of θ are restricted. Clearly, if $t_n(x_1, \cdots, x_n)$, $(n = 1, 2, \cdots$, ad inf.), is a uniformly consistent

estimate of θ, then also t_n^* is a uniformly consistent estimate of θ when $t_n^* = t_n$ when $a \leqq t_n \leqq b$, $t_n^* = a$ when $t_n < a$ and $t_n^* = b$ when $t_n > b$. Thus, without loss of generality, we can restrict ourselves to estimates t_n which can take values only in the interval $[a, b]$. Uniform consistency of t_n is then equivalent with the condition

$$\lim_{n=\infty} E[(t_n - \theta)^2 \mid \theta, \xi_1, \cdots, \xi_n] = 0 \tag{2.1}$$

uniformly in θ and the ξ's. For any chance variable u the symbol $E(u \mid \theta, \xi_1, \xi_2, \cdots)$ denotes the expected value of u when $\theta, \xi_1, \xi_2, \cdots$ are the true parameter values.

In [1] a non-negative function $W(t_n, \theta)$, called weight function, is introduced which expresses the loss suffered when t_n is the value of the estimate and θ is the true value of the parameter. The risk is defined in [1] as the expected value of the loss, i.e., the risk is given by

$$r_n(\theta, \xi_1, \cdots, \xi_n) = E[W(t_n, \theta) \mid \theta, \xi_1, \cdots, \xi_n]. \tag{2.2}$$

If we put $W(t_n, \theta) = (t_n - \theta)^2$, we have

$$r_n(\theta, \xi_1, \cdots, \xi_n) = E[(t_n - \theta)^2 \mid \theta, \xi_1, \cdots, \xi_n]. \tag{2.3}$$

It can easily be verified that Assumptions 1–4 in section 3 of [1] are fulfilled for the weight function $W(t_n, \theta) = (t_n - \theta)^2$.[3] Thus, all results obtained in [1] can be applied to the risk function given in (2.3). According to Theorem 4.1 in [1] the risk function given in (2.3) is a continuous function of $\theta, \xi_1, \cdots, \xi_n$ for any arbitrary estimate t_n. We shall denote the maximum of (2.3) with respect to $\theta, \xi_1, \cdots, \xi_n$ by $r_n[t_n]$. Thus $r_n[t_n]$ is a functional which associates a non-negative value with any estimate function t_n.

It follows from (2.1) that t_n is a uniformly consistent estimate of θ if and only if

$$\lim_{n=\infty} r_n[t_n] = 0 \tag{2.4}$$

For any θ and for any n let $F_n(\xi_1, \cdots, \xi_n \mid \theta)$ be a cumulative distribution function of $\xi_1, \cdots, \xi_n$. Let, furthermore,

$$\begin{aligned} &q_n(x_1, \cdots, x_n \mid \theta, F_n) \\ &\quad = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_n(x_1, \cdots, x_n \mid \theta, \xi_1 \cdots, \xi_n)\, dF_n(\xi_1, \cdots, \xi_n \mid \theta). \end{aligned} \tag{2.5}$$

We do not require that $F_1, F_2, \cdots$, etc. satisfy the consistency relations, i.e., $\lim_{\xi_{n+1}=\infty} F_{n+1}(\xi_1, \cdots, \xi_{n+1} \mid \theta)$ is not necessarily equal to $F_n(\xi_1, \cdots, \xi_n \mid \theta)$.

[3] In verifying Assumption 4, we may assume that p_n is always > 0, since for any given values $\theta, \xi_1, \cdots, \xi_n$ we may restrict the domain of $(x_1, \cdots, x_n)$ to the subset of the sample space where $p_n > 0$

Hence, also the distributions q_n do not necessarily satisfy the consistency relations. Clearly

$$(2.6) \qquad r_n[t_n] \geqq \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (t_n - \theta)^2 q_n(x_1, \cdots, x_n \mid \theta, F_n)\, dx_1, \cdots, dx_n$$

for any θ and any F_n. Hence, (2.4) and (2.6) imply that if t_n is a uniformly consistent estimate of θ, then t_n remains a uniformly consistent estimate of θ also when q_n is the distribution of $X_1, \cdots, X_n$ for any arbitrary choice of F_n.

For each n let $C_n(\theta, \xi_1, \cdots, \xi_n)$ be a joint cumulative distribution function of $\theta, \xi_1, \cdots, \xi_n$. If this is regarded as an a priori distribution of $\theta, \xi_1, \cdots, \xi_n$, and if our aim is to choose t_n so that

$$(2.7) \qquad \begin{aligned} &E(t_n - \theta)^2 \\ &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} (t_n - \theta)^2 p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)\, dC_n\, dx_1 \cdots dx_n \end{aligned}$$

is a minimum, then the best choice of t_n is to put it equal to the a posteriori mean value of θ. Let $t_n^*(x_1, \cdots, x_n; C_n)$ denote the a posteriori mean value of θ when C_n is the a priori distribution, i.e.,

$$(2.8) \qquad t_n^*(x_1, \cdots, x_n, C_n) = \frac{\int \theta p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)\, dC_n}{\int p_n(x_1, \cdots, x_n \mid \theta, \xi_1 \cdots, \xi_n)\, dC_n}$$

where the integration is to be taken over the whole domain of the parameters $\theta, \xi_1, \cdots, \xi_n$. Let, furthermore, $\bar{r}_n[C_n]$ denote the value of (2.7) when $t_n = t_n^*(x_1, \cdots, x_n; C_n)$. According to Theorem 4.4 in [1] there exists a particular distribution C_n^0, called a least favorable distribution, such that

$$(2.9) \qquad \bar{r}_n[C_n] \leqq \bar{r}_n[C_n^0]$$

for all C_n. Let

$$(2.10) \qquad t_n^0(x_1, \cdots, x_n) = t_n^*(x_1, \cdots, x_n; C_n^0).$$

It follows from Theorems (4.5) and (5.1) in [1] that for any estimate t_n we have

$$(2.11) \qquad r_n[t_n] \geqq r_n[t_n^0] = \bar{r}_n(C_n^0).$$

Hence, a necessary and sufficient condition for the existence of a uniformly consistent estimate of θ is that

$$(2.12) \qquad \lim_{n=\infty} \bar{r}_n[C_n^0] = 0$$

Let $F_n(\xi_1, \cdots, \xi_n \mid \theta)$ denote the conditional cumulative distribution of $\xi_1, \cdots, \xi_n$ for given θ that results from the joint distribution $C_n(\theta, \xi_1, \cdots \xi_n)$ and let $F_n^0(\xi_1, \cdots, \xi_n \mid \theta)$ correspond to $C_n^0(\theta, \xi_1, \cdots, \xi_n)$. Clearly, any uniformly consistent estimate of θ with respect to $p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)$

is a uniformly consistent estimate also with respect to $q_n(x_1, \cdots, x_n \mid \theta, F_n)$ for any F_n. On the other hand, if $q_n(x_1, \cdots, x_n \mid \theta, F_n^0)$ admits a uniformly consistent estimate of θ, equation (2.12) must hold and, therefore, $p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)$ admits a uniformly consistent estimate of θ. Hence we arrive at the following theorem:

THEOREM 2.1. *A necessary and sufficient condition that*

$$p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)$$

admit a uniformly consistent estimate of θ is that $q_n(x_1, \cdots, x_n \mid \theta, F_n)$ admit a uniformly consistent estimate of θ for any arbitrary choice of F_n.

3. Amount of information contained in the first n observations concerning the parameter θ. We shall make the following assumptions:

Assumption 1. The first two derivatives of $p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)$ with respect to θ exist.

Assumption 2. We have

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \operatorname*{Max}_{\theta} \left| \frac{\partial p_n}{\partial \theta} \right| dx_1 \cdots dx_n < \infty \tag{3.1}$$

and

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \operatorname*{Max}_{\theta} \left| \frac{\partial^2 p_n}{\partial \theta^2} \right| dx_1 \cdots dx_n < \infty \tag{3.2}$$

for any n.

Assumption 3. The integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial^2 \log q_n(x_1, \cdots, x_n \mid \theta, F_n)}{\partial \theta^2} q_n(x_1, \cdots, x_n \mid \theta, F_n)\, dx_1 \cdots dx_n$$

exists for any θ, F_n and n where q_n is defined by (2.5).

Since

$$\frac{\partial^2 \log q_n}{\partial \theta^2} = \frac{1}{q_n} \frac{\partial^2 q_n}{\partial \theta^2} - \left(\frac{\partial \log q_n}{\partial \theta} \right)^2$$

and since, because of Assumptions 1 and 2,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial^2 q_n}{\partial \theta^2} dx_1 \cdots dx_n = 0,$$

we have

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial^2 \log q_n}{\partial \theta^2} q_n\, dx_1 \cdots dx_n = - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{+\infty} \left(\frac{\partial \log q_n}{\partial \theta} \right)^2 q_n\, dx_1 \cdots dx_n.$$

Let

$$(3.4)\quad c_n(\theta) = \operatorname*{g\,l\,b}_{F_n} \left\{ - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\frac{\partial^2 \log q_n}{\partial \theta^2} \right) q_n \, dx_1 \cdots dx_n \right\}$$

Clearly $c_n(\theta) \geqq 0$. We shall now show that

$$(3.5)\qquad c_{n+1}(\theta) \geqq c_n(\theta) \qquad \text{for } n = 1, 2, \cdots, \text{ad inf.}$$

In fact, we can write

$$(3.6)\qquad \frac{-\partial^2 \log q_{n+1}(x_1, \cdots, x_{n+1} \mid \theta, F_{n+1})}{\partial \theta^2} = - \frac{\partial^2 \log q_n(x_1, \cdots, x_n \mid \theta, F_n^*)}{\partial \theta^2} - \frac{\partial^2 \log f_{n+1}(x_{n+1} \mid x_1, \cdots, x_n, \theta, F_{n+1})}{\partial \theta^2}$$

where $F_n^* = \lim_{\xi_{n+1} = \infty} F_{n+1}(\xi_1, \cdots, \xi_{n+1} \mid \theta)$ and $f_{n+1}(x_{n+1} \mid x_1, \cdots, x_n, \theta, F_{n+1})$ is the conditional probability density function of X_{n+1} given the values of $x_1, \cdots, x_n$ and assuming that the joint density function of $X_1, \cdots, X_{n+1}$ is given by $q_{n+1}(x_1, \cdots, x_{n+1} \mid \theta, F_{n+1})$ Since $c_n(\theta) \leqq$ expected value of

$$- \frac{\partial^2 \log q_n(x_1, \cdots, x_n \mid \theta, F_n^*)}{\partial \theta^2}$$

and since the expected value of $- \dfrac{\partial^2 \log f_{n+1}}{\partial \theta^2}$ is $\geqq 0$, inequality (3.5) must hold.

In analogy with R. A. Fisher's information function, we shall call $c_n(\theta)$ the amount of information contained in the first n observations regarding θ. We shall now prove the following theorem:

THEOREM 3.1. *If* $\lim_{n=\infty} c_n(\theta) \leqq c < \infty$ *over a finite non-degenerate θ-interval I, then there is no uniformly consistent estimate of θ.*

PROOF. If for any n, $c_n(\theta) \leqq c < \infty$ over the interval I, for each n there exists a distribution $F_n(\xi_1, \cdots, \xi_n \mid \theta)$ such that

$$(3.7)\qquad 0 \leqq - \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\partial^2 \log q_n(x_1, \cdots, x_n \mid \theta, F_n)}{\partial \theta^2} \cdot q_n(x_1, \cdots, x_n \mid \theta, F_n) \, dx_1 \cdots dx_n \leqq c + 1$$

for all n and for all θ in I. Let t_n be any estimate and let

$$(3.8)\qquad \begin{aligned} b_n(\theta) = E(t_n - \theta) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (t_n - \theta) q_n(x_1, \cdots, x_n \mid \theta, F_n) \, dx_1 \cdots dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} t_n q_n(x_1, \cdots, x_n \mid \theta, F_n) \, dx_1 \cdots dx_n - \theta. \end{aligned}$$

Since t_n is bounded, it follows from Assumptions 1 and 2 that $\dfrac{db_n(\theta)}{d\theta}$ exists and is

a continuous function of θ. According to a theorem by Cramér [2] we have

$$(3.9)\qquad E(t_n - \theta)^2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (t - \theta)^2 q_n \, dx_1 \cdots dx_n \geqq \frac{\left(1 + \frac{db_n}{\partial\theta}\right)^2}{c + 1}$$

for all θ in I. Thus, in order that $\lim_{n=\infty} E(t_n - \theta)^2 = 0$ uniformly in θ, we must have

$$(3.10)\qquad \lim_{n=\infty} \frac{db_n(\theta)}{d\theta} = -1$$

uniformly in θ over I. Let I be the interval ranging from g to h ($g < h$). From (3.10) it follows that

$$(3.11)\qquad \lim_{n=\infty} [b_n(h) - b_n(g)] = g - h.$$

Hence

$$\liminf_{n=\infty} \max_{\theta \text{ in } I} [b_n(\theta)]^2 \geqq \frac{(g-h)^2}{4}.$$

Since $E(t_n - \theta)^2 \geqq [b_n(\theta)]^2$, $E(t_n - \theta)^2$ cannot converge to zero uniformly in θ and Theorem 3.1 is proved.

4. Formula for $c_n(\theta)$ when $p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)$ is equal to $\varphi_1(x_1 \mid \theta, \xi_1)\, \varphi_2(x_2 \mid \theta, \xi_2) \cdots \varphi_n(x_n \mid \theta, \xi_n)$. Let $g_i(x_i \mid x_1, \cdots, x_{i-1}, \theta, F_n)$ be the conditional probability density of X_i given $x_1, \cdots, x_{i-1}$ when the joint density function of $x_1, \cdots, x_n$ is given by $q_n(x_1, \cdots, x_n \mid \theta, F_n)$, ($i \leqq n$). Clearly,

$$(4.1)\qquad -E\left(\frac{\partial^2 \log q_n}{\partial\theta^2}\right) = -\sum_{i=1}^{n} E\left(\frac{\partial^2 \log g_i}{\partial\theta^2}\right).$$

Now

$$(4.2)\qquad g_i(x_i \mid x_1, \cdots, x_{i-1}, \theta, F_n) = \int_{-\infty}^{\infty} \varphi_i(x_i \mid \theta, \xi_i)\, dH_i(\xi_i \mid x_1, \cdots, x_{i-1}, \theta, F_n)$$

where $H_i(\xi_i \mid x_1, \cdots, x_{i-1}, \theta, F_n)$ denotes the conditional cumulative distribution of ξ_i given $x_1, \cdots, x_{i-1}$, assuming that $F_n(\xi_1, \cdots, \xi_n \mid \theta)$ is the joint cumulative distribution of $\xi_1, \cdots, \xi_n$ and $p_n(x_1, \cdots, x_n \mid \theta, \xi_1, \cdots, \xi_n)$ is the joint density of $X_1, \cdots, X_n$ for any given values of $\theta, \xi_1, \cdots, \xi_n$.

It follows from (4.2) that

$$-\int_{-\infty}^{+\infty} \frac{\partial^2 \log g_i}{\partial\theta^2}\, g_i \, dx_i \geqq c_{ni}(\theta)$$

$$= \operatorname*{g.l.b.}_{G_i(\xi_i)} \left\{ -\int_{-\infty}^{+\infty} \left[\frac{\partial^2 \log \int_{-\infty}^{+\infty} \varphi_i(x_i \mid \theta, \xi_i)\, dG_i(\xi_i)}{\partial\theta^2} \int_{-\infty}^{\infty} \varphi_i \, dG_i \right] dx_i \right\}$$

where $C_i(\xi_i)$ may be any cumulative distribution of ξ_i. Hence

$$\text{g.l.b.}_{F_n} \left[-E \left(\frac{\partial^2 \log g_i}{\partial \theta^2} \right) \right] = c_{ni}(\theta) \tag{4.3}$$

and, therefore,

$$c_n(\theta) = \sum_{i=1}^{n} c_{ni}(\theta). \tag{4.4}$$

The quantity $c_{ni}(\theta)$ is simply the amount of information contained in the ith observation alone. Thus, formula (4.4) says that if $X_1, \cdots, X_n$ are independent, the total information contained in the first n observations is equal to the sum of the amounts of information contained in each of these observations singly.

REFERENCES

[1] A. Wald, "Statistical decision functions which minimize the maximum risk", *Annals of Math*, Vol. 46 (1945)
[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ Press, 1946.

INVERSION FORMULAE FOR THE DISTRIBUTION OF RATIOS

By John Gurland

University of California, Berkeley

1. Summary. The use of the repeated Cauchy principal value affords greater facility in the application of inversion formulae involving characteristic functions. Formula (2) below is especially useful in obtaining the inversion formula (1) for the distribution of the ratio of linear combinations of random variables which may be correlated. Formulae (1), (10), (12) generalize the special cases considered by Cramer [2], Curtiss [4], Geary [6], and are free of some restrictions they impose. The results are further generalized in section 6, where inversion formulae are given for the joint distribution of several ratios. In section 7, the joint distribution of several ratios of quadratic forms in random variables $X_1, X_2, \cdots, X_n$ having a multivariate normal distribution is considered

2. Introduction. We shall write

$$\oint\oint \cdots \oint g(t_1, t_2, \cdots, t_n)\, dt_1\, dt_2 \cdots dt_n$$

$$= \lim_{\substack{\epsilon_i \to 0 \\ T_i \to \infty}} \iint \cdots \int_{\epsilon_i < |t_i| < T_i} g(t_1, t_2, \cdots, t_n)\, dt_1\, dt_2 \cdots dt_n,$$

which might be called the repeated Cauchy principal value of

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(t_1, t_2, \cdots, t_n)\, dt_1 \cdots dt_n,$$

and which we shall use frequently. The results of this article may be regarded as extensions of the following theorem proved in section 4.

Theorem 1. *Let $X_1, X_2, \cdots, X_n$ have the joint distribution function $F(x_1, x_2, \cdots, x_n)$ with corresponding characteristic function $\phi(t_1, t_2, \cdots, t_n)$. Let $G(x)$ be the distribution function of $(a_1X_1 + \cdots + a_nX_n)/(b_1X_1 + \cdots + b_nX_n)$, where $a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n$ are real numbers. If*

$$P\left\{\sum_1^n b_j x_j \leq 0\right\} = 0,$$

then

$$G(x) + G(x - 0) = 1 - \frac{1}{\pi i} \oint \frac{\varphi\{t(a_1 - b_1 x), \cdots, t(a_n - b_n x)\}}{t}\, dt, \tag{1}$$

3. An inversion formula for distribution functions. Let $F(x)$ be a distribution function and $\phi(t)$ be the corresponding characteristic function. Then the following inversion formula holds:

(2) $$F(\xi) + F(\xi - 0) = 1 - \frac{1}{\pi i} \oint e^{-it\xi} \phi(t) \frac{dt}{t}.$$

PROOF.

$$\left(\int_{-T}^{-\epsilon} + \int_{\epsilon}^{T}\right) e^{-it\xi} \phi(t) \frac{dt}{t} = \left(\int_{-T}^{-\epsilon} + \int_{\epsilon}^{T}\right) \frac{e^{-it\xi}}{t} dt \int_{-\infty}^{\infty} e^{itx}\, dF(x)$$
$$= \int_{-\infty}^{\infty} dF(x) \left(\int_{-T}^{-\epsilon} + \int_{\epsilon}^{T}\right) e^{it(x-\xi)} \frac{dt}{t},$$

by the Fubini theorem on the inversion of integrals. But

$$\frac{1}{\pi i} \oint e^{it(x-\xi)} \frac{dt}{t} = \operatorname{sgn}(x - \xi),$$

where $\operatorname{sgn} y = -1, 0, 1$ according as $y < 0, y = 0, y > 0$. Since $\int_{-T}^{T} \frac{\sin at}{t} dt$ is uniformly bounded in T, the principle of bounded convergence for Lebesgue integrals implies that

$$\frac{1}{\pi i} \lim_{\substack{\epsilon \to 0 \\ T \to \infty}} \int_{-\infty}^{\infty} dF(x) \left(\int_{-T}^{-\epsilon} + \int_{\epsilon}^{T}\right) e^{it(x-\xi)} \frac{dt}{t} = \int_{-\infty}^{\infty} \operatorname{sgn}(x - \xi)\, dF(x)$$
$$= \left(\int_{-\infty}^{\xi - 0} + \int_{(\xi)} + \int_{\xi+0}^{\infty}\right) \operatorname{sgn}(x - \xi)\, dF(x)$$
$$= -F(\xi - 0) + 1 - F(\xi).$$

The required result follows at once.

Another form of (2) may be obtained as follows: Let $H(x)$, $K(x)$ be distribution functions, and $\psi(t)$, $\chi(t)$ the corresponding characteristic functions. Setting $F = H, \phi = \psi, \xi = 0$ in (2) yields

$$H(0) + H(0 - 0) = 1 - \frac{1}{\pi i} \oint \psi(t) \frac{dt}{t},$$

while setting $F = K, \phi = \chi$, in (2) yields

$$K(\xi) + K(\xi - 0) = 1 - \frac{1}{\pi i} \oint \chi(t) e^{-it\xi} \frac{dt}{t}.$$

Clearly

(3) $$K(\xi) + K(\xi - 0) = H(0) + H(0 - 0) + \frac{1}{\pi i} \oint \frac{\psi(t) - \chi(t) e^{-it\xi}}{t} dt.$$

If $H = K$, then $\psi = \chi$, and (3) reduces to a well-known inversion formula (cf Kendall [7, p. 91]).

4. Distribution of the ratio $(a_1 X_1 + \cdots + a_n X_n)/(b_1 X_1 + \cdots + b_n X_n)$ with denominator positive. THEOREM 1 *Let $X_1, X_2, \cdots, X_n$ have the joint*

distribution function $F(x_1, x_2, \cdots, x_n)$ *with corresponding characteristic function* $\phi(t_1, t_2, \cdots, t_n)$. *Let* $G(x)$ *be the distribution function of* $(a_1X_1 + \cdots + a_nX_n)/(b_1X_1 + \cdots + b_nX_n)$ *where* $a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n$ *are real numbers. If* $P\{\Sigma_1^n b_jX_j \leq 0\} = 0$, *then*

$$G(x) + G(x - 0) = 1 - \frac{1}{\pi i} \oint \frac{\phi\{t(a_1 - b_1x), \cdots, t(a_n - b_nx)\}}{t} dt.$$

PROOF.[9] Note that

$$P\left\{\frac{\Sigma a_iX_i}{\Sigma b_iX_i} \leq x\right\} = P\{\Sigma(a_i - b_ix)X_i \leq 0\},$$

and let $R_x(\xi) = P\{\Sigma(a_i - b_ix)X_i \leq \xi\}$ and $\chi_x(t)$ be the corresponding characteristic function. Clearly $R_x(0) = G(x)$ and

$$\chi_x(t) = \phi\{t(a_1 - b_1x), \cdots, t(a_n - b_nx)\}.$$

On applying (2) to $R_x(\xi)$ and setting $\xi = 0$, the required result follows at once. If (3) is applied in place of (2), with $K = G$, we obtain

$$G(x) + G(x - 0) = H(0) + H(0 - 0)$$

$$+ \frac{1}{\pi i} \oint \frac{\psi(t) - \phi\{t(a_1 - b_1x), \cdots, t(a_n - b_nx)\}}{t} dt. \tag{4}$$

We shall consider (3) and (4) when $n = 2$ and

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Two cases will be treated separately; first, when X_1, X_2 are independent, second, when X_1, X_2 may be correlated.

If X_1, X_2 are independent, and $F(x_1, \infty) = F_1(x_1)$, $F(\infty, x_2) = F_2(x_2)$, with corresponding characteristic functions $\phi_1(t)$, $\phi_2(t)$ then (1) becomes

$$G(x) + G(x - 0) = 1 - \frac{1}{\pi i} \oint \frac{\phi_1(t)\phi_2(-tx)}{t} dt, \tag{5}$$

while (4) becomes, taking $H = F$

$$G(x) + G(x - 0) = \frac{1}{\pi i} \oint \frac{\phi_2(t) - \phi_1(t)\phi_2(-tx)}{t} dt. \tag{6}$$

Cramér [2, p. 46] proves, for X_1, X_2 independent and $F_2(0) = 0$, that

$$G(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\phi_2(t) - \phi_1(t)\phi_2(-tx)}{t} dt$$

under the following conditions:

(i) X_1 and X_2 have finite means;

(ii) $\int_1^\infty \left|\frac{\phi_2(t)}{t}\right| dt < \infty.$

If X_1, X_2 may be correlated, then (1) becomes

$$G(x) + G(x - 0) = 1 - \frac{1}{\pi i} \oint \frac{\phi(t, -tx)}{t}\, dt; \tag{7}$$

while (4) becomes, taking $H = F$,

$$G(x) + G(x - 0) = \frac{1}{\pi i} \oint \frac{\phi_2(t) - \phi(t, -tx)}{t}\, dt. \tag{8}$$

Professor P. L. Hsu, in a course of lectures attended by the author at the Statistical Laboratory, University of California, gave the following result of Cramér, which was stated thus, using the above notation:

$$G(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\phi_2(t) - \chi_x(t)}{t}\, dt, \tag{9}$$

provided $\int_{-\infty}^{\infty} \left| \frac{\phi_2(t) - \chi_x(t)}{t} \right| dt < \infty$, where $F_2(0) = 0$ and $\chi_x(t)$ is defined above expression (4).

The following corollary is obtained from (1) according to well-known theorems concerning differentiation under the integral sign:

COROLLARY. *Suppose $\phi(t_1, t_2, \cdots, t_n)$ is the characteristic function corresponding to $X_1, X_2, \cdots, X_n$, and $G(x)$ is the distribution function of*

$$(a_1X_1 + \cdots + a_nX_n)/(b_1X_1 + \cdots + b_nX_n);$$

then, if $P\{\Sigma b_iX_i \leq 0\} = 0$,

$$G'(x) = \frac{1}{2\pi i} \oint \left[\sum_{k=1}^{n} b_k \frac{\partial \phi(t_1, \cdots, t_n)}{\partial t_k} \right]_{t_k = t(a_k - b_k x)} dt, \tag{10}$$

in every interval in which the integral converges uniformly.

If $n = 2$, and

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

then

$$G'(x) = \frac{1}{2\pi i} \oint \left[\frac{\partial \phi(t_1, t_2)}{\partial t_2} \right]_{t_2 = -t_1 x} dt_1 . \tag{11}$$

Cramér [3, p. 317, exercise 6] states the following result:

If $F(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f(u, v)\, du\, dv$, and $F_2(0) = 0$, then

$$G'(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \left[\frac{\partial \phi(t_1, t_2)}{\partial t_2} \right]_{t_2 = -t_1 x} dt_1 ,$$

if the integral is uniformly convergent with respect to x.

Geary [6] has shown that if $F(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f(u, v)\, du\, dv$, $F_2(0) = 0$, and $\lambda(t, v) = \int_{-\infty}^{\infty} e^{itv} f(u, v)\, du$, then

$$G'(x) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\left[\frac{\partial\phi(t_1, t_2)}{\partial t_2}\right]_{t_2=-t_1x} dt_1,$$

provided

(i) $$\phi(t_1\ t_2) = 0 \text{ for } t_2 = \pm\infty,$$

(ii) $$\int_0^{\infty} dy \int_{-\infty}^{\infty} y\lambda(t, y)e^{-ityx}\, dt = \int_{-\infty}^{\infty} dt \int_0^{\infty} y\lambda(t, y)e^{-ityx}\, dy.$$

Formula (1) can be employed in the case $n = 2$, X_1, X_2 are independent, and

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

to obtain closed expressions for the distribution functions of ratios in which the variable in the numerator and that in the denominator may have any one of the following four distributions: Binomial, Rectangular, χ^2, Normal. In the case of the four ratios with the binomially distributed variable as the denominator, a translation must be made to ensure positiveness of the denominator. For the four ratios with the normally distributed variable as denominator, the distribution function obtained is approximate; and the approximation is good if $P\{X_2 \leq 0\}$ is sufficiently small (cf. Geary [5])

5. Distribution of the ratio $(a_1X_1 + \cdots + a_nX_n)/(b_1X_1 + \cdots + b_nX_n)$, with denominator positive or negative. The following theorem will be proven:

THEOREM 2. *Let $G(x)$ be the distribution function of $(a_1X_1 + \cdots + a_nX_n)/(b_1X_1 + \cdots + b_nX_n)$ where $a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_n$ are real numbers. If $P\{\Sigma_1^n b_1X_1 = 0\} = 0$, then*

(12) $$G(x) + G(x - 0) = 1 - \frac{1}{\pi i}\oint \frac{\phi^+\{t(a_1 - b_1x), \cdots, t(a_n - b_nx)\} + \phi^-\{t(a_1 - b_1x), \cdots, t(a_n - b_nx)\}}{t}\, dt,$$

where

$$\phi^+(t_1, t_2, \cdots, t_n) = \iint_{\Sigma b_k x_k > 0} \cdots \int e^{i(t_1x_1+\cdots+t_nx_n)}\, dF(x_1, x_2, \cdots, x_n),$$

$$\phi^-(t_1, t_2, \cdots, t_n) = \iint_{\Sigma b_k x_k < 0} \cdots \int e^{i(t_1x_1+\cdots+t_nx_n)}\, dF(x_1, x_2, \cdots, x_n)$$

PROOF. Let $R_x(\xi) = P\{\Sigma b_k X_k > 0\}\cdot P\{\Sigma X_k(a_k - b_k x) \le \xi \mid \Sigma b_k X_k > 0\}$

$$+ P\{\Sigma b_k X_k < 0\}\ P\{\Sigma X_k(a_k - b_k x \ge - \xi \mid \Sigma b_k X_k < 0\}.$$

Then $R_x(\infty) = 1$, $R_x(-\infty) = 0$, and $R_x(\xi)$ is non-decreasing in ξ and continuous on the right. Hence $R_x(\xi)$ is a distribution function (Cramér [2, p. 11]) It can be shown by a proof analogous to that used by Curtiss [4] that the characteristic function of $R_x(\xi)$ is

$$\phi^+\{t(a_1 - b_1 x), \cdots, t(a_n - b_n x)\} + \phi^-\{t(b_1 x - a_1), \cdots, t(b_n x - a_n)\}$$

Since $R_x(0) = G(x)$, application of (2) to $R_x(\xi)$ yields the required result.

6. Inversion formulae for multidimensional distribution functions. The n-dimensional analogue of (2) will now be given, and will be applied to obtain inversion formulae for the joint distribution of several ratios

Let $X_1, X_2, \cdots, X_n$ have the joint distribution function $F(x_1, x_2, \cdots, x_n)$ and the corresponding characteristic function $\phi(t_1, t_2, \cdots, t_n)$. Let

$$\phi_{j_1, j_2, \cdots, j_k}(t_1, t_2, \cdots, t_k)$$

be the characteristic function corresponding to the marginal joint distribution function of $X_{j_1}, X_{j_2}, \cdots, X_{j_k}$, where the set $j_1, j_2, \cdots, j_k$ is a permutation of k of the integers $1, 2, \cdots, n$. Note that

$$\phi(t_1, t_2, \cdots, t_n) = \phi_{1,2,\cdots,n}(t_1, t_2, \cdots, t_n).$$

The summation $\sum_{(i_1, i_2, \cdots, i_n)} F(\xi_{1i_1}, \xi_{2i_2}, \cdots, \xi_{ni_n})$, which will appear below is to be interpreted as follows:

Defining $\xi_{ji_j} = \xi_j$ if $i_j = 1$,

$\qquad\qquad = \xi_j - 0$ if $i_j = 0$,

then $\sum_{(i_1, i_2, \cdots, i_n)} F(\xi_{1i_1}, \cdots, \xi_{ni_n})$ will mean that the summation is to be taken over all binary numbers $.i_1 i_2 \cdots i_n$.

Using the notation of the preceding paragraph, we can state the following theorem.

THEOREM 3. *Let $A_0, A_1, \cdots, A_n$ satisfy the $n + 1$ equations*

$$\sum_{k=0}^{n-r-1} \binom{n-r}{k} A_{r+k} = 1, \qquad A_n = -1, \qquad (r = 0, 1, 2, \cdots, n-1),$$

where $\binom{m}{p}$ as usual, denotes the binomial coefficient.

Then

$$(13)\qquad (-1)^{n+1} \sum_{(i_1, i_2, \cdots, i_n)} F(\xi_{1i_1}, \cdots, \xi_{ni_n}) = A_0 + \sum_{k=1}^{n} \frac{A_k}{(\pi i)^k} \sum_{j_1 < j_2 < \cdots < j_k} \oint \oint \cdots \oint$$
$$\cdot \exp\{-i(t_1 \xi_{j_1} + \cdots + t_k \xi_{j_k})\} \phi_{j_1 j_2 \cdots j_k}(t_1, t_2, \cdots, t_k) \frac{dt_1\, dt_2 \cdots dt_k}{t_1 t_2 \cdots t_k}.$$

PROOF: Since the theorem is already proved for $n = 1$ (section 3), and since

$$\frac{1}{(\pi i)^n} \oint \oint \cdots \oint e^{-i(t_1\xi_1+\cdots+t_n\xi_n)} \phi(t_1, t_2, \cdots, t_n) \frac{dt_1\, dt_2 \cdots dt_n}{t_1 t_2 \cdots t_n}$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \operatorname{sgn}(x_1 - \xi_1) \operatorname{sgn}(x_2 - \xi_2) \cdots$$

$$\cdots \operatorname{sgn}(x_n - \xi_n)\, dF(x_1, x_2, \cdots, x_n),$$

the theorem could be proven by induction. The result is obtained more quickly, however, by noting that it suffices to consider the case of independent $X_1, X_2, \cdots, X_n$.

It may be remarked that if $(\xi_1, \xi_2, \cdots, \xi_n)$ is a continuity point of $F(x_1, x_2, \cdots, x_n)$, the left-hand member of (13) becomes

$$(-1)^{n+1} 2^n F(\xi_1, \xi_2, \cdots, \xi_n),$$

and also that differentiation of (13) yields

$$\frac{\partial^n F(\xi_1, \xi_2, \cdots, \xi_n)}{\partial \xi_1 \partial \xi_2 \cdots \partial \xi_n} = \left(\frac{1}{2\pi}\right)^n \oint \oint \cdots \oint e^{-i(t_1\xi_1+\cdots+t_n\xi_n)} \phi(t_1, t_2, \cdots, t_n)\, dt_1\, dt_2 \cdots dt_n, \tag{14}$$

in every n-dimensional interval in which the integral converges uniformly. This agrees with well-known results concerning Fourier inversion formulae.

An inversion formula for the joint distribution of p ratios

$$\frac{a_{1i} X_1 + a_{2i} X_2 + \cdots + a_{ni} X_n}{b_{1i} X_1 + b_{2i} X_2 + \cdots + b_{ni} X_n}; \qquad i = 1, 2, \cdots, p \qquad (1 \leq p \leq n),$$

can be obtained from (13) by a method similar to that applied in section 4. The following theorem holds:

THEOREM 4. *Let*

$$G(\xi_1, \xi_2, \cdots, \xi_p) = P\left\{\frac{\Sigma a_{j1} X_j}{\Sigma b_{j1} X_j} \leq \xi_1; \cdots, \frac{\Sigma a_{jp} X_j}{\Sigma b_{jp} X_j} \leq \xi_p\right\}$$

and $\phi(t_1, t_2, \cdots, t_n)$ *be the characteristic function corresponding to* $X_1, X_2, \cdots, X_n$. *Then, if* $P\{\Sigma b_{jk} X_j \leq 0\} = 0$ $(k = 1, 2, \cdots p)$,

$$(-1)^{p+1} \sum_{(i_1, i_2, \cdots i_p)} G(\xi_{1i_1}, \xi_{2i_2}, \cdots, \xi_{pi_p})$$

$$= A_0 + \sum_{k=1}^{p} \frac{A_k}{(\pi i)^k} \sum_{j_1 < j_2 < \cdots < j_k} \oint \oint \cdots \oint \phi\left\{\sum_{l=1}^{k} t_l(a_{1j_l} - b_{1j_l} \xi_{j_l}), \right. \tag{15}$$

$$\left. \cdots, \sum_{l=1}^{k} t_l(a_{nj_l} - b_{nj_l} \xi_{j_l})\right\} \frac{dt_1\, dt_2 \cdots dt_k}{t_1 t_2 \cdots t_k}.$$

The following corollary is a generalization of (10) and follows by differentiation of (15).

COROLLARY. *Suppose $G(x_1, x_2, \cdots, x_p)$ is the joint distribution function of the p ratios*

$$\frac{a_{1j} X_1 + \cdots + a_{nj} X_n}{b_{1j} X_1 + \cdots + b_{nj} X_n},$$

and $\phi(t_1, t_2, \cdots, t_n)$ is the characteristic function corresponding to $X_1, X_2, \cdots, X_n$, then, if $P\{\sum_{i=1}^{n} b_{ij} X_i \leq 0\} = 0, j = 1, 2, \cdots p$.

$$\frac{\partial^p G(\xi_1 \xi_2 \cdots \xi_p)}{\partial \xi_1 \partial \xi_2 \cdots \partial \xi_p} = \left(\frac{1}{2\pi i}\right)^p \oint \oint \cdots \oint \tag{16}$$
$$\cdot \left[\sum_{k=1}^{n} b_{k1} b_{k2} \cdots b_{kp} \frac{\partial^p \phi(t_1, t_2, \cdots t_n)}{\partial t_k^p}\right]_{t_k = \sum_{j=1}^{p} \tau_j (a_{kj} - b_{kj} \xi_j)} d\tau_1 \, d\tau_2 \cdots d\tau_p,$$

in every p-dimensional interval in which the integral converges uniformly

7. Joint distribution of ratios of quadratic forms. *Let* $X_1, X_2, \cdots, X_n$ have the joint probability density function

$$f(x) = \frac{(\det B)^{\frac{1}{2}}}{(2\pi)^{n/2}} e^{-\frac{1}{2} x B x'}$$

where $x = (x_1, x_2, \cdots, x_n)$ and B is a positive definite symmetric matrix. Suppose Q is a positive semi-definite symmetric matrix of rank $r \leq n$ and $L_1, L_2, \cdots, L_p$ is a set of symmetric matrices. We wish to obtain the joint distribution function $G(\xi_1, \xi_2, \cdots, \xi_p)$ of the p ratios

$$\frac{XL_1X'}{XQX'}, \quad \frac{XL_2X'}{XQX'}, \cdots, \frac{XL_pX'}{XQX'},$$

where $X = (X_1, X_2, \cdots, X_n)$

The existence of such an orthogonal matrix S that $SQS' = I^{(r)}$, where $I^{(r)}$ is the diagonal matrix having the first r diagonal elements equal to unity and the rest equal to zero, is well-known. Let $X = YS$, $C = SBS'$, $M_i = SL_iS'$, and note that C and the M_i are symmetric matrices Also

$$G(\xi_1, \xi_2, \cdots, \xi_p) = P\left\{\frac{YM_1Y'}{YI^{(r)}Y'} \leq \xi_1, \cdots, \frac{YM_pY'}{YI^{(r)}Y'} \leq \xi_p\right\},$$

where $Y = (Y_1, Y_2, \cdots, Y_n)$ has the probability density function

$$g(y) = \frac{(\det C)^{\frac{1}{2}}}{(2\pi)^{n/2}} e^{-\frac{1}{2} y C y'},$$

and $y = (y_1, y_2, \cdots, y_n)$.

Suppose the L_i mutually commute in pairs Then so do the M_i; for $M_iM_j = SL_iS'SL_jS' = SL_iL_jS' = SL_jL_iS' = SL_jS'SL_iS' = M_jM_i$, since S is orthog-

onal. Hence, there is an orthogonal matrix U which simultaneously reduces each M to diagonal form; that is, $N = UMU'$ is a diagonal matrix (cf. Weyl [8, p. 25]).

Let $Y = ZU, D = UCU'$, so that

$$ZN_i Z' = \sum_{j=1}^{n} \mu_{ji} Z_j^2$$

$$G(\xi_1, \xi_2, \cdots, \xi_p) = P\left\{ \frac{\Sigma \mu_{j1} Z_j^2}{\Sigma \nu_j^{(r)} Z_j^2} \leq \xi_1 ; \cdots ; \frac{\Sigma \mu_{jp} Z_j^2}{\Sigma \nu_j^{(r)} Z_j^2} \leq \xi_p \right\},$$

where $\nu_j^{(r)} = 1$ if $j \leq r$;

$= 0$ if $j > r$,

and $Z = (Z_1, Z_2, \cdots, Z_n)$ has the probability density function

$$h(z) = \frac{(\det D)^{\frac{1}{2}}}{(2\pi)^{n/2}} e^{-\frac{1}{2}zDz'},$$

where $z = (z_1, z_2, \cdots, z_n)$.

We can now apply the results of section 6. If $\psi(t_1, t_2, \cdots, t_n)$ is the characteristic function corresponding to the joint distribution function of $Z_1^2, Z_2^2, \cdots, Z_n^2$ it is clear that

$$\psi(t_1, t_2, \cdots, t_n) = \left[\frac{\det D}{\det(D - 2iT)}\right]^{\frac{1}{2}},$$

where T is the diagonal matrix whose diagonal elements are $t_1, t_2, \cdots, t_n$. Applying (15), with $\phi = \psi$, we obtain, since G is obviously a continuous distribution function

$$
\begin{aligned}
&(-1)^{p+1} 2^p G(\xi_1, \xi_2, \cdots, \xi_p) \\
(17) \qquad &= A_0 + \sum_{k=1}^{p} \frac{A_k}{(\pi i)^k} \sum_{j_1 < j_2 < \cdots < j_k} \oint \oint \cdots \oint \psi \left\{ \sum_{l=1}^{k} w_l(\mu_{1j_l} - \nu_1^{(r)} \xi_{j_l}), \right. \\
&\qquad \left. \cdots, \sum_{l=1}^{k} w_l(\mu_{nj_l} - \nu_n^{(r)} \xi_{j_l}) \right\} \frac{dw_1 \cdots dw_k}{w_1 w_2 \cdots w_k} \\
&= A_0 + \sum_{k=1}^{p} \frac{A_k}{(\pi i)^k} \sum_{j_1 < j_2 < \cdots < j_k} \oint \oint \cdots \oint \\
&\left[\frac{\det D}{\det\left\{ d_{\alpha\beta} - 2i\delta_{\alpha\beta} \sum_{l=1}^{k} w_l(\mu_{\beta j_l} - \nu_\beta^{(r)} \xi_{j_l}) \right\}}\right]^{\frac{1}{2}} \frac{dw_1 dw_2 \cdots dw_k}{w_1 w_2 \cdots w_k},
\end{aligned}
$$

where $D = [d_{\alpha\beta}]$ and $\delta_{\alpha\beta}$ is the Kronecker delta.

It is, of course, evident that a result analogous to (17) could be obtained, by considering p ratios

$$\frac{XL_1X'}{XQ_1X'}, \quad \frac{XL_2X'}{XQ_2X'}, \cdots, \frac{XL_pX'}{XQ_pX'},$$

where the $2p$ matrices $L_1, L_2, \cdots, L_p, Q_1, Q_2, \cdots, Q_p$ are symmetric and mutually commute in pairs, and $Q_1, Q_2, \cdots, Q_p$ are positive semi-definite.

In the case $p = 1$ in (17) and for special classes of matrices L_1, Q_1, B the calculus of residues may be employed to obtain closed expressions for the distribution of

$$\frac{XL_1X'}{XQ_1X'}.$$

Formula (17) can be applied to obtain the joint distribution of serial correlation coefficients with different lags. The author plans to incorporate these results with those mentioned at the end of section 4 in a forthcoming paper, written jointly with Roy B. Leipnik.

The author wishes to acknowledge the valuable criticism of Professor H. Lewy, and especially the constructive advice and suggestions of Roy B Leipnik.

REFERENCES

[1] M. Bôcher, *Introduction to Higher Algebra*, Macmillan Co , New York, 1907.
[2] H. Cramér, "Random variables and probability distributions", *Cambridge Tracts in Mathematics*, No. 36, Cambridge, 1937.
[3] H Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.
[4] J. H. Curtiss, "Distribution of a quotient", *Annals of Math Stat* , Vol 12 (1941).
[5] R. C. Geary, "The frequency distribution of the quotient of two normal variates", *Roy. Stat. Soc. Jour* , Vol. 103 (1930)
[6] R. C Geary, "Extension of a theorem by Harald Cramér", *Roy. Stat Soc. Jour.*, Vol. 107 (1944).
[7] M G Kendall, *The Advanced Theory of Statistics*, Griffin Co , London, 1943.
[8] H Weyl, *The Theory of Groups and Quantum Mechanics*, Methuen and Co , Ltd , London, 1931

THE FACTORIAL APPROACH TO THE WEIGHING PROBLEM[1]

By O. Kempthorne

Iowa State College

1. Summary. The weighing problem is discussed from the point of view of factorial experimentation. The paper contains a brief description of the fractional replication of the 2^n factorial system. It is shown that optimum designs for the weighing problem may easily be obtained with this approach. This approach is valuable in indicating the structure of weighing problem designs, and the limited conditions under which such designs can give results of value.

2. Introduction. Considerable attention has been given recently to the problem of weighing a number of light objects on a scale [1, 2, 3, 6] The problem was originally proposed by Yates in his paper on complex experiments [4] as an example of a factorial experiment in which interactions between the factors tested would not be expected to exist. that is, the weight of say two objects could be assumed to be the sum of the weights of the objects weighed separately, after taking account of any necessary zero corrections Such a situation is comparatively rare in biological research when, for example, the effect on yield of a particular crop from the joint application of two nutrients is usually different from the sum of the effects of separate applications. In recent years attention has been given to the use of fractional replication in factorial experiments [7, 8, 9] and it is proposed in this paper to consider the weighing problem from this point of view.

3. The 2^n factorial system. A full description of the 2^n factorial system was given by Yates in his technical communication *The Design and Analysis of Factorial Experiments* [5]. Yates was particularly concerned with the analysis of such experiments and with the evolution of systems of confounding in order to reduce the number of plots in each block. The following brief account is given in order to facilitate the discussion of the weighing problem.

In a single replication of the 2^n system all combinations of n factors each at two levels are tested. With three factors, a, b, c, for example, the following eight combinations are tested: (1) a, b, ab, c, ac, bc, and abc, where (1) denotes the control, a the application of treatment a only, ab the application of treatments a and b, and so on. A set of seven independent comparisons between the eight test results is given formally by the expansion of the formula

$$\tfrac{1}{4}(a \pm 1)(b \pm 1)(c \pm 1),$$

[1] Journal Paper no J 1548 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project No 890

where at least one of the signs is negative. If, for instance the first sign only is taken to be negative, a formal expansion gives the expression

$$\tfrac{1}{4}\{abc - bc + ab - b + ac - c + a - 1\},$$

and this contrast of the observations gives the effect of the factor a averaged over the presence and absence of the factors b and c, which is denoted by effect A Similarly taking the negative sign in the second bracket only, we get the average effect B,

$$B = \tfrac{1}{4}\{abc - ac + ab - a + bc - c + b - 1\}.$$

Taking negative signs in the first and second brackets we obtain the interaction AB ·

$$AB = \tfrac{1}{4}\{abc + c + ab + (1) - ac - bc - a - b\}, \quad \text{and so on.}$$

The definition of effects and interactions may be presented very simply in geometrical terminology, by representing the treatment combinations as points of an n-dimensional lattice, each axis of the lattice having two points at unit distance apart. The control treatment will have coordinates $(0, 0, 0, \cdots, \cdots, 0)$, the treatment consisting of a only will have co-ordinates $(1, 0, 0, \cdots, 0)$ and so on. The effect A is then the difference of the mean yield of the treatments corresponding to the points lying on the hyperplane

$$x_1 = 0,$$

and the mean yield of those represented by points lying on the hyperplane

$$x_1 = 1.$$

The interaction of two factors a and b, represented by the axes x_1 and x_2 respectively, will be obtained from the difference of the mean yields of those plots for which

$$x_1 + x_2 = 0, \qquad \text{or } x_1 + x_2 = 2,$$

and those for which

$$x_1 + x_2 = 1.$$

The extensions to the above for three-factor and higher order interactions are simple. The interaction of factors a, b, and c, which are represented by coordinate axes x_1, x_2, and x_3, is given by the difference between the mean of plots represented by points for which

$$x_1 + x_2 + x_3 = 0 \text{ or } 2,$$

and those represented by points for which

$$x_1 + x_2 + x_3 = 1 \text{ or } 3;$$

in other words, it is the difference of the mean yields of those plots for which

$$x_1 + x_2 + x_3 = 0 \pmod 2$$

and of those for which

$$x_1 + x_2 + x_3 = 1 \pmod 2.$$

Each effect or interaction is then defined as the mean difference of two sets of plots, each set being represented by points on parallel hyperplanes, and the planes of one set of parallel hyperplanes lying between the planes of the other set. It is necessary to specify only the direction cosines of the hyperplanes in order to specify the effect or interaction, and the usual terminology for effects and interactions follows, in that the interaction of factors a, b, c, for example may be represented by the symbol ABC.

In the same way as effects and interactions are defined in terms of the yields of the several treatment combinations, the expected yield from each treatment combination may be expressed in terms of the mean level of yield and the true effects and interactions If the full set of combinations of the factorial scheme is tested, the best estimate of each true effect and interaction is the same function of the observed yields that the true effect or interaction is of the true yields. This fact is one of the advantages which follow from the use of the full factorial scheme

We are not concerned here with factorial experiments in which the factors have more than two levels, but when the number of levels of each factor is the same prime number, effects and interactions may be represented by products of powers of the symbols for the factors. In the case of two factors (a, b) at three levels, for example, the main effects may be represented by A, B, and the interactions by AB and AB^2, each symbol referring to the two independent contrasts between three sets, each of three plots

As an example of the use of the above representation, we may consider confounding, that is, the arrangement of the treatment combinations in blocks in order to reduce the experimental error. Suitable arrangements are such that contrasts between the blocks represent high order interactions which the experimenter is confident will be of negligible size.

If treatment combinations for which

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0 \pmod 2)$$

and for which

$$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = 0 \pmod 2)$$

are arranged in a particular block, then the coordinates of the treatment combinations in this block also lie on the hyperplane

$$(\alpha_1 + \beta_1)x_1 + (\alpha_2 + \beta_2)x_2 + \cdots + (\alpha_n + \beta_n)x_n = 0 \pmod 2),$$

where the coefficients $(\alpha_1 + \beta_1)$ must be reduced modulo two. If, therefore, the treatments are arranged in blocks so that two comparisons are block contrasts, then the generalised interaction of these contrasts is also a block contrast

4. Fractional replication. The principle of fractional replication follows very simply. Suppose only those treatment combinations whose yields all occur either in the positive or the negative part of a particular contrast are represented in the experiment, that is only those combinations represented by the points of the lattice for which say

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0 \pmod 2.$$

Then the comparison between the yields of those plots represented by

$$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = 0 \pmod 2$$

and by

$$\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n = 1 \pmod 2$$

will be identical with the comparison between the yields of plots represented by

$$(\alpha_1 + \beta_1)x_1 + (\alpha_2 + \beta_2)x_2 + \cdots + (\alpha_n + \beta_n)x_n = 0 \pmod 2$$

and by

$$(\alpha_1 + \beta_1)x_1 + (\alpha_2 + \beta_2)x_2 + \cdots + (\alpha_n + \beta_n)x_n = 1 \pmod 2.$$

The former of these two comparisons may be represented by the symbol $x_1^{\beta_1}x_2^{\beta_2} \cdots x_n^{\beta_n}$, and the latter by $x_1^{\alpha_1+\beta_1}x_2^{\alpha_2+\beta_2} \cdots x_n^{\alpha_n+\beta_n}$, where $x_1, \cdots, x_n$ are no longer coordinates but symbols for the n factors, which satisfy the relations, $x_i^\alpha = 1$, if $\alpha = 0 \pmod 2$. The equivalence of the two comparisons may be obtained by the use of an identity relationship in the symbols for the factors

$$I = x_1^{\alpha_1}x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$

where I is interpreted as unity, and only those combinations whose coordinates $(x_1, x_2, \cdots, x_n)$ satisfy one of the equations

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0, \qquad \text{or} = 1 \pmod 2,$$

are represented in the experiment. If this identity relationship is multiplied by the symbol $x_1^{\beta_1}x_2^{\beta_2} \cdots x_n^{\beta_n}$ by ordinary commutative algebra, reducing the powers modulo 2 where necessary, we obtain

$$x_1^{\beta_1}x_2^{\beta_2} \cdots x_n^{\beta_n} = x_1^{(\alpha_1+\beta_1)}x_2^{(\alpha_2+\beta_2)} \cdots x_n^{(\alpha_n+\beta_n)}.$$

It is more convenient to revert to the common use of capital letters A, B, C, etc. for effects corresponding to small letters a, b, c, etc. for the factors tested An experiment in half-replicate is then represented formally by an equation of the type

$$I = A^\alpha B^\beta C^\gamma \cdots$$

In such an experiment on n factors only 2^{n-1} treatment combinations will be tested. Of the $2^n - 1$ independent comparisons in a fully replicated experi-

ment, information on one comparison is lost completely since only those treatments which appear in the comparison with the same sign are represented: the remaining $2^n - 2$ independent comparisons of a fully replicated experiment are identical in pairs giving $2^{n-1} - 1$ independent comparisons. Each comparison is then said to have two aliases and measures the sum (or difference, depending on which half of the treatment combinations are used) of two effects, an effect and an interaction, or two interactions.

A quarter-replicated experiment can by the same process be represented by an identity relationship of the form

$$I = A^{\alpha_1}B^{\beta_1}C^{\gamma_1} \cdots = A^{\alpha_2}B^{\beta_2}C^{\gamma_2} \cdots = A^{(\alpha_1+\alpha_2)}B^{(\beta_1+\beta_2)}C^{(\gamma_1+\gamma_2)} \cdots$$

It is useful in the evolution of fractional designs to note that the elements in the identity relationship form an Abelian group.

Fractionally replicated experiments are formally identical with confounded experiments in that block differences may be regarded as additional factors in the confounded experiment. A 2^n experiment arranged in 2^p blocks, for example, may be regarded as a 1 in 2^p design of a 2^{n+p} experiment. Considerable care needs to be exercised in the use of fractionally replicated designs, but they have been found to be very useful in agricultural and biological research.

5. The weighing problem. The problem of weighing a number of objects may be regarded as the problem of the estimation of the effects of a number of factors which do not interact. To take a simple case, consider the estimation of the effects of factors a, b, and c for which one complete replicate would consist of the combinations

$$(1)\ a,\ b,\ ab,\ c,\ ac,\ bc, \text{ and } abc.$$

Suppose a half replicate design is used, based on the identity relationship

$$I = ABC.$$

The combinations tested would then consist of either the set $\{a, b, c, abc\}$ or the set $\{(1), ab, ac, bc\}$. If the former set were chosen, the comparison estimating the effect A could also be ascribed to the interaction BC, that estimating effect B also to the interaction AB, and that estimating effect C to the interaction AC, as can be observed by multiplying the identity relationship by A, B, and C in turn. If the experimenter is confident that the two-factor interactions are negligible, then any effect given by each comparison would be ascribed to the main effect.

6. Discussion of a particular case. We give the derivation of a design for weighing a particular number of objects, say ten Let the objects be denoted by a, b, c, d, e, f, g, h, k, l. Then the total number of combinations which could be tested is 2^{10}, that is 1024, but as we are confident that interactions are negligible, it is necessary only to estimate main effects.

A fractionally replicated design must consist of a number 2^p of combinations and this will be a 1 in 2^{10-p} design. A suitable fractionally replicated design consisting of 16 combinations will exist if it is possible to evolve an identity

relationship for a 1 in 64 design, such that each term in the relationship involves at least three letters A possible identity relationship for such a design contains the numbers of the Abelian group obtained from all combinations of the elements $1, ABC, CDE, EFG, GHK, ADL$, and AFH, with the rule that the square of each letter is to be equated to unity. Each possible comparison may then be due to any of the 64 effects or interactions which may be derived from this identity relationship. In other words, each comparison has 64 aliases: in the case of ten of the comparisons, only one of the aliases is a main effect, and for the remaining five comparisons the aliases are all interactions of at least two factors. The actual design may be written down by finding combinations of the letters which have the same number of letters in common with the unit element and the six three-factor interactions. These are themselves a group consisting of all combinations of unity and four combinations of letters. The sixteen combinations with an even number of letters in common with all the members of the identity group are found to be the following:

(1)	*abdef*,	*acefl*,	*bcdl*,
abfgkl,	*degkl*,	*bcegk*,	*acdfgk*,
fgh,	*abdegh*,	*aceghl*,	*bcdefghl*,
abhkl,	*defhkl*,	*bcefhk*,	*acdhk*

The estimation of effects from the results of the sixteen weighings is particularly easy, the weight of object a will be one-eighth of the difference between those weighings containing a and those not containing a. There are ten such contrasts which estimate the effects, and the remaining five contrasts may be used to obtain an estimate of the experimental error. If σ^2 is the variance of each weighing, the variance of the weight of a, that is, the effect A will be $(1/8 + 1/8)\sigma^2 = (1/4)\sigma^2$. The precision can be increased fourfold in the weighing problem with a chemical balance by interpreting the absence of each letter as the placing of the object in the left hand pan and the presence as the placing of the object in the right hand pan. Each effect will then measure twice the weight of the corresponding object and the estimated weight of each object will have a variance of $\sigma^2/16$, that is, the same precision as if each had been weighed by itself eight times in each pan, or sixteen times in all.

7. General case. The rules by which fractional designs may be constructed have been exemplified above and the procedure is simple, though laborious in the case of a large number of objects. It does not, therefore, seem worth while to enumerate particular designs for the weighing of particular numbers of objects. A general procedure in considering the design for a particular problem is as follows. Taking the case of a number n of objects, the experimenter should form a rough idea of the order of magnitude of the experimental error, σ say, and decide what accuracy he requires for his estimates of the weights, a standard error say of s. Then if he weighs 2^p combinations of the objects, the standard error of the estimate of each weight will be $2^{-p/2}\sigma$ in the case of the chemical balance This serves to determine $2^{p/2}$ and therefore p, and it is then necessary

to design a 2^n experiment of fraction 2^{n-p}. Alternatively, a design of higher fraction which can provide estimates may be replicated the corresponding number of times. In the case of the spring balance the corresponding standard error is $2^{-(p-1)/2}\sigma$ necessitating a design of higher fraction.

Designs of the type described above have some useful properties:

(1) the design automatically takes care of any bias in the balance,
(2) the effects or weights may be computed easily as indicated above,
(3) the effects or weights are uncorrelated,
(4) all the effects are measured with the same precision, and
(5) an estimate of the experimental error which is independent of the effects may be computed from the results

In considering the use for a particular problem of a design like the one discussed, it is important to understand completely the structure of the design. Such designs may well have real value for the weighing problem, but it is not easy to visualize other problems for which they would not give results capable of various interpretations. The use of the above designs depends on an assumption that interactions between pairs of factors are negligible, and this is generally not the case, for example, in biological research work, in which the experimenter may well be confident that interactions between three or more factors are negligible. In the particular case described in detail, there are only fifteen independent comparisons between the sixteen results which will be obtained, and it follows from the identity relationship that the comparison giving the effect A, also measures the two factor interactions BC, DL, and FH. If therefore the factor a has no effect and there is an interaction between factors b and c or the other two pairs of factors, the experimenter will conclude that the factor a has an effect. It is clear that under these circumstances the experimental results are worthless.

8. Efficiency of designs. In [2], Mood states for optimum designs such that when N weighings are made, the variance of the estimates of the weights are of the order of $\frac{\sigma}{\sqrt{N}}$ in the case of the chemical balance and $\frac{2\sigma}{\sqrt{N}}$ in the case of the spring balance, where σ^2 is the variance of a single weighing. As indicated above, when N is a power of 2, the fractional factorial designs result in the same variances. These designs are similar to those proposed by Kishen [6].

Mood dealt with the case in which the number of weighings was restricted, for example to be equal to the number of objects, and defined a best design as that which gave the smallest confidence region in the p-dimensional space for the estimates of the p-weights.

To take an example for the weighings of 3 objects with a spring balance with no bias he suggests the following design:

$$X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

where the rows of the array refer to weighing operations and the columns refer

to objects. If the results of the weighings are y_1, y_2, and y_3 respectively, the estimates of the weights b_1, b_2, and b_3 are given by the equation

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ -1/2 & 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

If σ^2 is the variance of a single weighing, then the variance of each estimate will be $[(1/2)^2 + (1/2)^2 + (-1/2)^2]\sigma^2 = (3/4)\sigma^2$: or if N $(= 0(\text{mod } 3))$ weighings are made by replicating the above system $N/3$ times, the variance of each estimate will be $9\sigma^2/4N$. The covariance of any two estimates is $(-1/4)\sigma^2$ so that the square of the correlation between any two estimates is $-1/9$. The fractional factorial design will yield estimates which have a somewhat higher variance, namely $4\sigma^2/N$. This increase in precision obtained in Mood's design has been obtained at the expense of obtaining correlated estimates which in addition are subject to any bias which the measuring instrument may have. It is doubted whether the use of such designs for any practical problem can be justified.

It is of interest to note that the concept of fractional replication may be extended to give designs requiring a number of weighings other than a power or two. For the weighing of 3 objects for example, a factorial design of fraction 3/4 could be used: it could consist of a half-replicate based on the identity $I = ABC$, and a quarter replicate based on the identity

$$I = A = BC = ABC.$$

There is, however, little point in discussing such designs for the weighing problem, as their efficiency as measured by the total number of weighings needed to achieve a particular degree of accuracy is lower than for the designs described in this paper.

REFERENCES

[1] HAROLD HOTELLING, "Some improvements in weighing and other experimental Techniques", *Annals of Math Stat*, Vol 15 (1944), pp 297–306.

[2] A M. MOOD, "On Hotelling's weighing problem", *Annals of Math Stat.*, Vol. 17 (1946), pp 432–446

[3] R. L PLACKETT AND J P. BURMAN, "The design of optimum multifactorial experiments", *Biometrika*, Vol. 33 (1946), pp 305–325.

[4] F. YATES, "Complex experiments", *Roy Stat Soc. Jour. Suppl.*, Vol 2 (1935), pp. 181–247.

[5] F YATES, "The design and analysis of factorial experiments", *Technical Communication No.* 35, Imperial Bureau of Soil Science, Harpenden, Herts, England, 1937.

[6] K. KISHEN, "On the design of experiments for weighing", *Annals of Math Stat.*, Vol. 14 (1945), pp 294–301.

[7] D. J. FINNEY, "The fractional replication of factorial arrangements", *Annals of Eugenics*, Vol. 12 (1945), pp. 291–301.

[8] D. J FINNEY, "Recent developments in the design of field experiments III. Fractional replication", *Jour of Agric Sci.*, Vol 36 (1946), pp 184–191.

[9] O. KEMPTHORNE, "A simple approach to confounding and fractional replication in factorial experiments", *Biometrika*, Vol 34 (1947), pp. 255–272.

MULTIPLE SAMPLING FOR VARIABLES

By Jack Silber

Roosevelt College

Summary. A multiple (sequential) sampling scheme designed to test certain hypothesis is discussed with the following assumption: X is a random variable with density function $P(x)$ which is piecewise continuous and differentiable at its points of continuity. Formulae are derived for the probability of acceptance and rejection of the hypothesis and for the expected number of samples necessary for reaching a decision. These formulae are found to depend on the solution of a Fredholm Integral equation. Explicit solutions to the problem are obtained for the case when $P(x)$ is rectangular by reducing the fundamental integral equation to a set of differential-difference equations. Several examples are given.

1. Introduction. A multiple sampling scheme is here proposed which is based on cumulative sums of random variables. Bartky [1] has developed a theory of multiple sampling for attributes when the attribute can take only two values with probability (p) and ($1 - p$) respectively. Formulae are there derived for the probabilities of acceptance and rejection of the null hypothesis and for the expected amount of sampling necessary for reaching a decision. In this paper the same type of formulae are developed for the case of variable sampling where the underlying probability law for the variable is given by a piecewise continuous function for which derivatives exist at its points of continuity.

The whole theory of multiple sampling is closely related to Wald's [2] theory of sequential tests. The fundamental difference is that in the latter, probabilities of errors of the first and second kinds are assigned, and acceptance and rejection criteria derived therefrom, while in the former the problem is solved in reverse order. There the acceptance and rejection criteria are assigned, and probabilities of eventual acceptance and rejection derived. For different parameter values, these are the probabilities of making errors of the first and second kinds.

The problem presented here is similar to that given by Wald [3] in his paper on cumulative sums In the present paper we waive the restriction that the expected number of items necessary for termination of the cumulating process be given explicitly as an integer. Since the theory given here is from the point of view of multiple sampling, the language appropriate to that theory will be used.

2. The sampling scheme. Let X be a random variable with probability density function $P(x)$ which is piecewise continuous. One variate, say x_1, is selected and if $x_1 > b$, the hypothesis (for example the null hypothesis with respect to the mean) is accepted, and if $x_1 < a$, the hypothesis is rejected. If,

however, $a \leq x_1 \leq b$, another variate x_2 is selected In the latter case similar criteria with respect to $x_1 + x_2$ determine whether the hypothesis is to be accepted or this method of sampling continued. Or more formally, let

$$S_r = \sum_{i=1}^{r} x_i \qquad (r = 1, 2, 3, \cdots),$$

where the cumulative sums S_r are formed sequentially as follows for any integer r the cumulating process is terminated by acceptance of the hypothesis if $S_r > b$ and rejection if $S_r < a$, but, if $a \leq S_r \leq b$ another variate x_{r+1} is selected and the sum S_{r+1} formed. The acceptance and rejection criteria are then applied as above. No attempt is made here to indicate the choice of the acceptance and rejection criteria.

3. The probability of acceptance. If at the rth unit the hypothesis is neither accepted nor rejected, then it must be true that $a \leq S_r \leq b$. Let us denote the probability that this condition holds by

$$\int_a^b Y_r(S_r)\, dS_r, \tag{3.1}$$

where $Y_r(S_r)$ is the probability density function for S_r in the above described sampling scheme The probability density function for S_{r+1} would then be given by

$$Y_{r+1}(S_{r+1}) = \int_a^b Y_r(S_r)P(S_{r+1} - S_r)\, dS_r. \tag{3.2}$$

The probabilities of accepting or rejecting the hypothesis on the rth trial are respectively

$$\int_b^{+\infty} Y_r(S_r)\, dS_r, \qquad \int_{-\infty}^{a} Y_r(S_r)\, dS_r, \tag{3.3}$$

and therefore the probabilities for eventual acceptance or rejection are given by

$$P_A = \sum_{r=1}^{\infty} \int_b^{\infty} Y_r(S_r)\, dS_r, \qquad P_R = \sum_{r=1}^{\infty} \int_{-\infty}^{a} Y_r(S_r)\, dS_r. \tag{3.4}$$

The probability that $a \leq S_n \leq b$ cannot exceed the probability that $a \leq T_n = x_1 + x_2 + x_3 + \cdots + x_n \leq b$ on a single sample of n variates, that is Pr $(a \leq S_n \leq b) \leq$ Pr $(a \leq T_n \leq b)$. For distributions with positive variance, it can be shown that the right member of the above inequality approaches zero as $n \to 0$. Therefore, the process of sampling as outlined above will eventually lead to acceptance or rejection of the hypothesis. See Wald [3, p. 284] for a direct proof that the probability that the left member of the above inequality holds for $n = 1, 2, 3, \cdots$ is zero.

Consider the linear integral (Fredholm) equation

$$Y(x) = Y_1(x) + \lambda \int_a^b P(x - s)Y(s)\, ds, \tag{3.5}$$

where $Y_1(x) = P(x)$ and assume a solution of the form

$$Y(x) = Y_1(x) + \lambda Y_2(x) + \lambda^2 Y_3(x) + \cdots . \tag{3.6}$$

That solutions, in power series in λ, of the Fredholm equation exist when the kernel $P(x - s)$ and the function $Y_1(x)$ have finite discontinuities is well known and the theory has been expounded by several authors. (For example see Goursat [4].) If the power series in λ is substituted in the integral equation we obtain

$$\begin{aligned} Y_1(x) &+ \lambda Y_2(x) + \lambda^2 Y_3(x) + \cdots \\ &= Y_1(x) + \lambda \int_a^b [Y_1(s) + \lambda Y_2(s) + \lambda^2 Y_3(s) + \cdots] P(x - s)\, ds \\ &= Y_1(x) + \lambda \int_a^b Y_1(s)P(x - s)\, ds + \lambda^2 \int_a^b Y_2(s)P(x - s)\, ds \\ &\qquad + \lambda^3 \int_a^b Y_3(s)P(x - s)\, ds + \cdots . \end{aligned} \tag{3.7}$$

Equating coefficients of like powers of λ we see that

$$Y_r(x) = \int_a^b Y_{r-1}(s)P(x - s)\, ds, \qquad (r = 2, 3, \cdots). \tag{3.8}$$

This, however, is the probability distribution for S_r, $r = 2, 3, \cdots$ under our sampling scheme, and therefore from the equations,

$$Y(x) = \sum_{r=1}^{\infty} \lambda^{r-1} Y_r(x) = Y_1(x) + \lambda \int_a^b P(x - s)Y(s)\, ds, \tag{3.9}$$

we have that the probability of eventual acceptance for $\lambda = 1$, is

$$\sum_{r=1}^{\infty} \int_a^b Y_r(S_r)\, ds_r = \int_a^{\infty} Y(x)\, dx. \tag{3.10}$$

Thus our problem of finding a formula for the probability of eventual acceptance or rejection of the statistical hypothesis under the above sampling scheme reduces to that of finding a solution of a linear integral equation.

The argument in this section has, of course, been entirely formal. However from the general theory of integral equations we know that the series solution (3.6) converges uniformly for $\lambda < 1/M(b - a)$ where $P(x) \leq M$, since $P(x)$ is a probability density function. In equations (3.4) and (3.10) we give solutions for $\lambda = 1$ and of course we assume that $M(b - a) < 1$. Since (3.6) is uniformly convergent the interchanges of integration and summation in (3.10) and (4.3) in the following section are allowable.

4. The expected amount of sampling. Since

$$\int_a^b Y_{r-1}(S_{r-1})\, dS_{r-1} \tag{4.1}$$

is the probability that the rth sample will be reached, then the probability that on the rth sample, the hypothesis will be either accepted or rejected becomes

$$\int_a^b Y_{r-1}(S_{r-1})\, dS_{r-1} - \int_a^b Y_r(S_r)\, dS_r\,, \tag{4.2}$$

that is, the first term in this expression gives the probability that no terminating decision is made on the $(r - 1)$st sample and the second term gives the probability that a like decision is made on the rth sample. The difference of the two then gives the probability that a terminating decision (acceptance or rejection) will be made on the rth sample. The expected number of units sampled will therefore be

$$\begin{aligned} E &= 1 - \int_a^b P(x)\, dx + \sum_{r=2}^{\infty} r \left[\int_a^b Y_{r-1}(S_{r-1})\, dS_{r-1} - \int_a^b Y_r(S_r)\, dS_r \right] \\ &= 1 + \sum_{r=1}^{\infty} \int_a^b Y_r(S_r)\, dS_r = 1 + \int_a^b \sum_{r=1}^{\infty} Y_r(x)\, dx \\ &= 1 + \int_a^b Y(x)\, dx. \end{aligned} \tag{4.3}$$

Thus, the amount of sampling expected before a terminating decision is reached also depends upon the solution of the integral equation. We proceed to discuss the problem when $P(x)$ is given by a rectangular distribution.

5. Reduction to differential equations when P(x) is rectangular. Consider the integral equation

$$Y^*(z) = P^*(z) + \lambda \int_a^b P^*(z - t) Y^*(t)\, dt, \tag{5.1}$$

where

$$\begin{aligned} P^*(z) &= \frac{1}{2c}, \qquad -c \le z - \alpha \le +c, \\ &= 0, \qquad z - \alpha > c \quad \text{or} \quad z - \alpha < -c, \end{aligned} \tag{5.2}$$

and in the integral equation

$$a + \alpha - c < z < b + \alpha - c.$$

The parameter α is restricted to the values $-c \le \alpha \le c$ for the following reasons The rejection criterion a cannot be greater than $c + \alpha$ for, if so, rejection will be automatic on the first sample. Similarly the acceptance criterion b must be greater than $-c + \alpha$ for otherwise, acceptance would be automatic on the first

trial. If $\alpha > c$ then, rejection can never take place if it does not take place on the first trial for in this case all $z > 0$. Similarly, if $\alpha < -c$ then, acceptance can never take place if it does not take place on the first trial for in this case all $z < 0$. Furthermore, in obtaining solutions of the integral equation, we will take α to be ≥ 0 This inequality is no real restriction since solutions for negative α can be obtained by symmetry from the solutions for positive α.

If we let $x = z - \alpha$ then

$$Y^*(x + \alpha) = P^*(x + \alpha) + \lambda \int_a^b P^*(x + \alpha - t)Y^*(t)\,dt, \tag{5.3}$$

or

$$Y(x) = P(x) + \lambda \int_a^b P(x - t)Y^*(t)\,dt, \tag{5.4}$$

where

$$\begin{aligned} P(x) &= \frac{1}{2c}, & -c \leq x \leq +c; \\ &= 0, & x < -c \quad \text{or} \quad x > +c. \end{aligned} \tag{5.5}$$

Now let $s = t - \alpha$, then

$$\begin{aligned} Y(x) &= P(x) + \lambda \int_{a-\alpha}^{b-\alpha} P(x - \alpha - s)Y^*(s + \alpha)\,ds \\ &= P(x) + \lambda \int_{a-\alpha}^{b-\alpha} P(x - \alpha - s)Y(s)\,ds. \end{aligned} \tag{5.6}$$

We have thus transformed our equation to one in which $P(x)$ becomes symmetrical with respect to the line $x = 0$. Furthermore, the probability of acceptance becomes

$$P_A = \int_{b-\alpha}^{\infty} Y(x)\,dx, \tag{5.7}$$

and the expected amount of sampling becomes

$$E = 1 + \int_{a-\alpha}^{b-\alpha} Y(x)\,dx \tag{5.8}$$

Also, x now has the following range: $a - c < x < b + c$. If we now make the transformation $x - \alpha - s = y$, then

$$Y(x) = P(x) + \int_{x-b}^{x-a} P(y)Y(x - \alpha - y)\,dy, \tag{5.9}$$

and the following cases present themselves.

If $x - a < -c$ or $x - b > +x$, then $Y(x) \equiv P(x)$, since $P(y) \equiv 0$,

If $x - b < -c < x - a < +c$, then

$$Y(x) = P(x) + \frac{\lambda}{2c}\int_{-c}^{x-a} Y(x - \alpha - y)\,dy, \tag{5.10}$$

where

$$\begin{aligned} &a - c \le x \le a + c \qquad \text{when } b - a \ge 2c,\\ &a - c \le x \le b - c \qquad \text{when } b - a \le 2c. \end{aligned} \tag{5.11}$$

If $x - b < -c < +c < x - a$, then

$$Y(x) = P(x) + \frac{\lambda}{2c}\int_{-c}^{+c} Y(x - \alpha - y)\,dy, \tag{5.12}$$

where

$$a + c \le x \le b - c \text{ and } b - c \ge 2c. \tag{5.13}$$

If $-c < x - b < x - a < +c$, then

$$Y(x) = P(x) + \frac{\lambda}{2c}\int_{x-b}^{x-a} Y(x - \alpha - y)\,dy, \tag{5.14}$$

where

$$b - c \le x \le a + c \text{ and } b - a \le 2c. \tag{5.15}$$

If $-c < x - b < +c < x - a$, then

$$Y(x) = P(x) + \frac{\lambda}{2c}\int_{x-b}^{+c} Y(x - \alpha - y)\,dy, \tag{5.16}$$

where

$$\begin{aligned} &b - c \le x \le b + c \text{ when } b - a \ge 2c,\\ &a + c \le x \le b + c \text{ when } b - a \le 2c. \end{aligned} \tag{5.17}$$

Transforming back to the variable s, we have for the case $b - a \ge 2c$,

$$\begin{aligned} Y(x) &= P(x) + \frac{\lambda}{2c}\int_{a-\alpha}^{x-\alpha+c} Y(s)\,ds \quad \text{for} \quad a - c \le x \le a + c,\\ &= P(x) + \frac{\lambda}{2c}\int_{x-\alpha-c}^{x-\alpha+c} Y(s)\,ds \quad \text{for} \quad a + c \le x \le b - c,\\ &= P(x) + \frac{\lambda}{2c}\int_{x-\alpha-c}^{b-\alpha} Y(s)\,ds \quad \text{for} \quad b - c \le x \le b + c, \end{aligned} \tag{5.18}$$

and for the case $b - a \le 2c$,

$$\begin{aligned} Y(x) &= P(x) + \frac{\lambda}{2c}\int_{a-\alpha}^{x-\alpha+c} Y(s)\,ds \quad \text{for} \quad a - c \le x \le b - c,\\ &= P(x) + \frac{\lambda}{2c}\int_{a-\alpha}^{b-\alpha} Y(s)\,ds \quad \text{for} \quad b - c \le x \le a + c,\\ &= P(x) + \frac{\lambda}{2c}\int_{x-\alpha-c}^{b-\alpha} Y(s)\,ds \quad \text{for} \quad a + c \le x \le b + c \end{aligned} \tag{5.19}$$

In all of the above equations, the integral is a continuous function of x, α, a, b, c while $P(x)$ has a discontinuity at $x = +c$ and $x = -c$, the jump at these points being of amount $1/2c$. The function $Y(x)$ will therefore be such that

$$
\begin{aligned}
&Y(-c + 0) - Y(-c - 0) = 1/2c, \\
(5.20) \qquad & \\
&Y(c - 0) - Y(c + 0) = 1/2c.
\end{aligned}
$$

If we now differentiate the above sets of integral equations with respect to x we obtain the following sets of differential-difference equations for the case $\lambda = 1$. If $b - a \geq 2c$,

$$
\begin{aligned}
Y'(x) &= \frac{1}{2c} Y(x - \alpha + c) && \text{for} \quad a - c \leq x \leq a + c, \\
(5,21) \qquad &= \frac{1}{2c} \{Y(x - \alpha + c) - Y(x - \alpha - c)\} && \text{for} \quad a + c \leq x \leq b - c, \\
&= -\frac{1}{2c} Y(x - \alpha - c) && \text{for} \quad b - c \leq x \leq b + c,
\end{aligned}
$$

and, if $b - a \leq 2c$,

$$
\begin{aligned}
Y'(x) &= \frac{1}{2c} Y(x - \alpha + c) && \text{for} \quad a - c \leq x \leq b - c, \\
(5.22) \qquad &= 0 && \text{for} \quad b - c \leq x \leq a + c, \\
&= -\frac{1}{2c} Y(x - \alpha - c) && \text{for} \quad a + c \leq x \leq b + b,
\end{aligned}
$$

the derivatives holding for all points except at $x = -c$ and $x = +c$.

Although a technique has been devised to solve the above equations for finite a and b, mathematical difficulties of a computational character are encountered when $(b - a)$ is made large. Note that there are only three essential parameters in the above problem since c can be taken as the unit of measurement In the technique illustrated by the following examples, α has been fixed as has $(b - a)$, i e. the solutions shown in the examples below are general only insofar as one parameter is concerned The essential feature of the technique is that the range of $Y(x)$ has been further subdivided so as to make its points of discontinuity end points of subdivisions of its range, and thus $Y(x)$ becomes continuous from the right or left in every subinterval of its range.

6. Example I: $b - a = 2c$, $c = 1$, $\alpha = 0$. In this case x ranges from $(a - 1)$ or $(-c)$, whichever is smaller, to $(b + 1)$ or $(+c)$, whichever is larger. If $-c < a - 1$, then $Y(x) = P(x)$ for $-c \leq x < a - 1$, and if $b + 1 < +c$, then $Y(x) = P(x)$ for $b + 1 < x < -c$. For x between $a - 1$ and $b + 1$ the domain of the differential-difference equations is divided as follows, where a is now restricted to the interval $-1 \leqq a \leqq 0$.

$$
\begin{aligned}
(6.1)\quad Y_i'(x) &= \tfrac{1}{2}Y_{i+2}(x+1) \text{ where for } && i = 1, && a-1 \le x < -1,\\
&&& i = 2, && -1 \le x < a,\\
&&& i = 3, && a \le x < 0,\\
&&& i = 4, && 0 \le x < a+1;\\
Y_i'(x) &= -\tfrac{1}{2}Y_{i-2}(x-1) \text{ where for } && i = 5, && a+1 \le x < +1,\\
&&& i = 6, && +1 \le x < a+2,\\
&&& i = 7, && a+2 \le x < +2,\\
&&& i = 8, && +2 \le x < a+3.
\end{aligned}
$$

The above are the equations corresponding to (5.21) for the given example

Differentiating the equations for $i = 3, 4, 5, 6$ and making certain obvious substitutions we obtain the following second order differential equations,

$$
(6.2)\qquad Y_i''(x) = -\tfrac{1}{4}Y_i(x), \qquad i = 3, 4, 5, 6,
$$

where the domains for x are as in (6.1). If we solve the equations (6 2) and substitute in the remaining equations in (6.1) we obtain the following set of equations,

$$
\begin{aligned}
&Y_i(x) = A_{i+2}\sin\tfrac{1}{2}(x+1) - B_{i+2}\cos\tfrac{1}{2}(x+2) + K_i, && i = 1, 2,\\
(6.3)\qquad &Y_i(x) = A_i\cos\tfrac{1}{2}x + B_i\sin\tfrac{1}{2}x, && i = 3, 4, 5, 6,\\
&Y_i(x) = -A_{i-2}\sin\tfrac{1}{2}(x-1) + B_{i-2}\cos\tfrac{1}{2}(x-1) + K_i, && i = 7, 8,
\end{aligned}
$$

where again the domains are as in (6.1)

From continuity considerations we have the boundary conditions

$$
\begin{gathered}
Y_1(a-1) = Y_8(a+3) = 0, \qquad Y_1(-1) - \tfrac{1}{2} = Y_2(-1), \qquad Y_2(a) = Y_3(a),\\
Y_3(0) = Y_4(0), \qquad Y_4(a+1) = Y_5(a+1), \qquad Y_5(1) = Y_6(1) + \tfrac{1}{2},\\
Y_6(a+2) = Y_7(a+2), \qquad Y_7(2) = Y_8(2).
\end{gathered}
$$

These boundary conditions yield certain relationships between the constants The equations so determined, however, do not form a consistent set of linear equations in the A_i, B_i, $K_i \cdots$. If we integrate out the equations (5 18), sectionally, the following relationships between the constants are obtained.

$$
\begin{aligned}
&A_i = A_{i+2}\sin\tfrac{1}{2} - B_{i+2}\cos\tfrac{1}{2}, \qquad B_i = B_{i+2}\cos\tfrac{1}{2} + B_{i+2}\sin\tfrac{1}{2}, \qquad i = 3, 4,\\
&K_2 = -(A_4 - A_5)\sin\tfrac{1}{2}(a+1) - (B_5 - B_4)\cos\tfrac{1}{2}(a+1)\\
&\qquad\qquad = \tfrac{1}{2} + B_4 - B_3 + K_1,\\
&K_7 = A_3 - A_4 + K_8, \qquad K_8 = A_6\sin\tfrac{1}{2}(a+2) - B_6\cos\tfrac{1}{2}(a+2),\\
(6.4)\qquad&\\
&B_3 = \tfrac{1}{2} + B_4 + K_1 - A_4 + A_8 + (A_4 - A_5)\sin\tfrac{1}{2}(a+1)\\
&\qquad\qquad + (B_5 - B_4)\cos\tfrac{1}{2}(a+1),\\
&A_4 = A_3 + K_8 + (A_4 - A_5)\sin\tfrac{1}{2}(a+1) + (B_5 - B_4)\cos\tfrac{1}{2}(a+1),\\
&K_1 = -A_3\sin\frac{a}{2} + B_3\cos\frac{a}{2}.
\end{aligned}
$$

From these equations it is easily seen that $A_4 = A_3$ and $K_1 = K_2 = K_7 = K_8$. Furthermore, the following set of consistent linear equations is obtained, after several simple manipulations and substitutions.

$$
\begin{aligned}
&\left\{\sin \tfrac{1}{2}(a+2) + \sin\frac{a}{2}\cdot \sin\tfrac{1}{2}\right\} A_6 \\
&\qquad - \left\{\cos\frac{a}{2}\right\} B_3 + \left\{\cos\tfrac{1}{2}(a+2) + \sin\frac{a}{2}\cos\tfrac{1}{2}\right\} B_6 = 0, \\
&\left\{-\sin\tfrac{1}{2}(a+2) + \cos\tfrac{1}{2}(a+2) - \cos\frac{a}{2}\cdot\sin\tfrac{1}{2}\right\} A_6 + \left\{\sin\frac{a}{2}\right\} B_3 \\
&\qquad + \left\{\sin\tfrac{1}{2}(a+2) + \cos\tfrac{1}{2}(a+2) - \cos\frac{a}{2}\cdot\cos\tfrac{1}{2}\right\} B_6 = 0, \\
&\qquad \{\cos\tfrac{1}{2}\} A_6 - B_3 + \{\sin\tfrac{1}{2}\} B_6 = 0.
\end{aligned}
\tag{6.5}
$$

All the other constants can be obtained from the solutions for A_6, B_3, B_6 in (6.5) Letting Δ equal the determinant of coefficients in (6.5) and using the relationships (6.4) we obtain the following solutions:

$$
\begin{aligned}
\Delta &= 2 - 2\sin\tfrac{1}{2} - \cos\tfrac{1}{2}, \\
\Delta A_4 &= \tfrac{1}{2}\{\cos\tfrac{1}{2} - \cos a/2 \cdot \sin\tfrac{1}{2}(a+1)\} = \Delta A_3, \\
\Delta B_4 &= \tfrac{1}{2}\{\sin\tfrac{1}{2} - \sin a/2 \cdot \sin\tfrac{1}{2}(a+1) + \cos\tfrac{1}{2} - 1\}, \\
\Delta A_6 &= \tfrac{1}{2}\{\sin 1 - \cos\tfrac{1}{2} + \cos a/2 \cdot \cos\tfrac{1}{2}(a+2)\}, \\
\Delta B_6 &= \tfrac{1}{2}\{\sin\tfrac{1}{2}(a+2)\cos a/2 - \sin\tfrac{1}{2} - \cos 1\}, \\
\Delta B_3 &= \tfrac{1}{2}\{1 - \sin a/2 \cdot \sin\tfrac{1}{2}(a+1) - \sin\tfrac{1}{2}\}, \\
\Delta A_5 &= \tfrac{1}{2}\{\cos\tfrac{1}{2} - \sin^2\tfrac{1}{2}(a+1)\}, \\
\Delta B_5 &= \tfrac{1}{2}\{\sin\tfrac{1}{2} + \sin\tfrac{1}{2}(a+1)\cdot\cos\tfrac{1}{2}(a+1) - 1\}, \\
\Delta K_1 &= \tfrac{1}{2}\left\{\cos\frac{a}{2}\sin\tfrac{1}{2}(a+2)\right\} = \Delta K_2 = \Delta K_7 = \Delta K_8.
\end{aligned}
\tag{6.6}
$$

If we now integrate $Y(x)$, equation (6.3) sectionally, i.e. from the left end point to the right end point of each sub-interval of its range and then add up appropriate areas, we obtain the following formulae for the probabilities of acceptance and rejection and for the expected amount of sampling:

$$
\begin{aligned}
P_R &= \frac{1}{\Delta}\{\cos\tfrac{1}{2}(a+1) + \sin a/2 - \cos a/2 + \Delta K_2\}, \\
P_A &= \frac{1}{\Delta}\{2 - \cos\tfrac{1}{2} - 2\sin\tfrac{1}{2} + \sin\tfrac{1}{2}(a+1) \\
&\qquad - \cos\tfrac{1}{2}(a+1) - \sin a/2 + \Delta K_2\}, \\
E &= \frac{1}{\Delta}\{\cos a/2 - 2\sin a/2 - \sin\tfrac{1}{2}(a+1)\}.
\end{aligned}
\tag{6.7}
$$

7. Example II: $a = 1, c = 3, b - a = 4$. In this case, as in the previous one, $Y(x) = P(x)$ for $-3 \leq x < a - 3$ when $a - c = a - 3 < -3$ and if $b + c = a + 7 < 3$ then $Y(x) = P(x)$, $a + 7 \leq x < 3$. For $a - 3 \leq x \leq a + 7$ where a takes on only integral values between -5 and 3, we have the following set of differential-difference equations:

$$
(7.1) \qquad
\begin{aligned}
Y'_{a+j}(x) &= \tfrac{1}{6}Y_{a+j+2}(x+2), & j &= -3, -2, -1, 0,\\
&= 0, & j &= 1, 2,\\
&= -\tfrac{1}{6}Y_{a+j-4}(x-4), & j &= 3, 4, 5, 6.
\end{aligned}
$$

If we integrate the above equations for $j = 1, 2$, substitute in the equations for $j = -1, 0, 5, 6$, integrate, and then substitute in the remaining equations, we obtain the solutions

$$
(7.2) \qquad
\begin{aligned}
Y_{a+j}(x) &= \tfrac{1}{72} A_{a+j+4}(x+2)^2 + \tfrac{1}{6}A_{a+j+2}\, x + A_{a-j}\,, & j &= -3, -2;\\
&= \tfrac{1}{6}A_{a+j+2}\, x + A_{a+j} & j &= -1, 0;\\
&= A_{a+j} & j &= 1, 2;\\
&= -\tfrac{1}{72}A_{a+j-2}(x-4)^2 - \tfrac{1}{6}A_{a+j-4}\, x + A_{a+j}\,, & j &= 3, 4;\\
&= -\tfrac{1}{6}A_{a+j-4}\, x + A_{a+j}\,, & j &= 5, 6
\end{aligned}
$$

As in the previous example we now use (5 22). Integrating out (5.22) sectionally, certain relationships between the A_{a+j}, $j = -3, -2, \cdots 6$, are obtained. These yield

$$
(7.3) \qquad
\begin{aligned}
A_{a+1} &= \tfrac{1}{28}\{12P_{a-1} + 12P_a + 39P_{a+1} + 9P_{a+2}\},\\
A_{a+2} &= \tfrac{1}{28}\{12P_{a-1} + 12P_a + 11P_{a+1} + 37P_{a+2}\},\\
A_{a-1} &= -\frac{a}{56}\{4P_{a-1} + 4P_a + 13P_{a+1} + 3P_{a+2}\}\\
&\qquad + \tfrac{1}{168}\{228P_{a-1} + 60P_a + 55P_{a+1} + 17P_{a+2}\},\\
A_a &= -\frac{a}{168}\{12P_{a-1} + 12P_a + 11P_{a+1} + 37P_{a+2}\}\\
&\qquad + \tfrac{1}{168}\{60P_{a-1} + 228P_a + 55P_{a+1} + 17P_{a+2}\},
\end{aligned}
$$

where P_{a+j} is the value of $P(x)$ for $a + j \leq x \leq a + j + 1, j = -3, \cdot\ 6$. All of the other constants can be found in terms of those given in equations (7.3). If we now integrate (7.2) sectionally and perform several simple manipulations, we arrive at the following formulas:

$$
(7.4) \qquad
\begin{aligned}
P_R &= \sum_{j=-0}^{-2} P_{a+j} + \frac{9a-5}{216}A_{a+1} + \frac{3a-1}{216}A_{a+2} + \frac{3}{12}A_{a-1} - \frac{1}{12}A_a\,,\\
P_A &= \sum_{j=3}^{9} P_{a+j} + \frac{3a-89}{216}A_{a+1} + \frac{9a-131}{216}A_{a+2} + \tfrac{1}{12}A_{a-1} + \tfrac{3}{12}A_a\,,\\
E &= 1 + \frac{2a-11}{12}A_{a+1} + \frac{2a-13}{12}A_{a+2} + A_{a+1} + A_a
\end{aligned}
$$

Although P_{a+j}, $j = -6, -5, -4, 7, 8, 9$, have not appeared in previous derivations in this example, they have been inserted in the above formulas to cover the cases in which $a - c > -c$ or $b + c < c$.

It should be mentioned that Kac [5] obtained the distribution of n (the expected amount of sampling) by a process similar to that given in this paper. It is also interesting to note that the present paper could have been written entirely in the language of problems in Random Walk.

The author has also worked on the case in which the distribution $P(x)$ is triangular and parabolic. In these, as in the case of the rectangular distribution discussed in this paper for $b - a$ large, the equivalent differential-difference equations are of large orders making the computation of solutions extremely tedious. As Kac [5] pointed out, the task of obtaining solutions in closed form for the case when $P(x)$ is the normal law is extremely difficult.

REFERENCES

[1] W. Bartky, "Multiple sampling with constant probability", *Annals of Math. Stat.*, Vol. 14 (1942), p 363.

[2] A. Wald, "Sequential tests of statistical hypotheses", *Annals of Math Stat.*, Vol. 16 (1945), p. 17.

[3] A. Wald, "On cumulative sums of random variables", *Annals of Math. Stat.*, Vol 15 (1944), pp 283–284

[4] E Goursat, *Cours d'Analyse Mathematique*, Gauthier, Paris, 1923, Vol 3.

[5] M. Kac, "Random walk in the presence of absorbing barriers", *Annals of Math. Stat.*, Vol 16 (1945), p. 62

ON THE CHARACTERISTIC FUNCTIONS OF THE DISTRIBUTIONS OF ESTIMATES OF VARIOUS DEVIATIONS IN SAMPLES FROM A NORMAL POPULATION

By M. Kac

Cornell University

1. Summary. An explicit formula for the characteristic function of the deviation

$$\frac{1}{n}\sum_{k=1}^{n} | X_k - \bar{X} |^{\alpha}, \qquad \alpha > 0,$$

is derived for samples from a normal population. For $\alpha = 1$ one can calculate the probability density function but the result does not seem to be in complete agreement with a recent formula of Goodwin [1].

2. Introduction. Let $X_1, X_2, \cdots, X_n$ be independent, normally distributed random variables each having mean 0 and variance 1.

Let, as usual,

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

and denote by $Y_n(\alpha)$ the deviation

$$Y_n(\alpha) = \frac{1}{n}\sum_{k=1}^{n} | X_k - \bar{X} |^{\alpha}, \qquad \alpha > 0 \tag{1}$$

The purpose of this note is to show that

$$\begin{aligned} F_n(\xi) &= E\{\exp (i\xi Y_n(\alpha))\} \\ &= \frac{1}{\sqrt{n}(\sqrt{2\pi})^{n+1}} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} e^{-x^2/2}\, e^{i/n(\xi|x|^{\alpha}+\eta x)}\, dx \right]^n d\eta. \end{aligned} \tag{2}$$

It is easy to check that for $\alpha = 2$ one obtains the well known expression

$$\left(1 - \frac{2i\xi}{n}\right)^{-(n-1)/2}$$

Moreover, if $\alpha = 1$ one can actually find the probability density of $Y_n(1)$ The resulting expression is fairly complicated and it strongly resembles an expression recently obtained by Goodwin [1]. Except for the relatively simple case $n = 3$, it does not seem easy to verify that our formula is equivalent to that of Goodwin.

Although deviations corresponding to values of α different from 1 and 2 are of little practical value the explicit formula (2) may be of some interest. It is also hoped that the method of derivation may prove useful in other cases.

3. The derivation of (2). We start with the observation that

$$\bar{X} \text{ and } Y_n(\alpha)$$

are statistically independent (see e.g. Daly [2]).

Denote by

$$E^*\{\,|\bar{X}| < \epsilon, \exp(i\xi Y_n(\alpha))\}$$

the integral of $\exp(i\xi Y_n(\alpha))$ extended over that portion of the sample space in which $|\bar{X}| < \epsilon$. Thus the conditional expectation $E\{\exp(i\xi Y_n(\alpha)) \mid |\bar{X}| < \epsilon\}$ is given by the formula

$$E\{\exp(i\xi X_n(\alpha)) \mid |\bar{X}| < \epsilon\} = \frac{E^*\{\,|\bar{X}| < \epsilon, \exp(i\xi Y_n(\alpha))\}}{\text{Prob}\{\,|\bar{X}| < \epsilon\}}.$$

Because of the independence of $\bar{X}$ and $Y_n(\alpha)$ we have

$$E\{\exp(i\xi Y_n(\alpha))\} = \frac{E^*\{\,|\bar{X}| < \epsilon, \exp(i\xi Y_n(\alpha))\}}{\text{Prob}\{\,|\bar{X}| < \epsilon\}}. \tag{3}$$

For the sake of simplicity we assume now that $\alpha \geq 1$ and note that

$$\left|\exp(i\xi Y_n(\alpha)) - \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right| \leq \frac{\xi}{n}\sum_1^n \big|\,|X_k|^\alpha - |X_k - \bar{X}|^\alpha\big|$$

$$\leq \frac{\alpha\xi|\bar{X}|}{n}\sum_1^n (|X_k| + |\bar{X}|)^{\alpha-1}.$$

Thus, on the portion of the sample space where $|\bar{X}| < \epsilon$, we have

$$\left|\exp(i\xi Y_n(\alpha)) - \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right| \leq \frac{\alpha\xi\epsilon}{n}\sum_1^n (|X_k| + \epsilon)^{\alpha-1}$$

and consequently

$$\left|E^*\{\,|\bar{X}| < \epsilon, \exp(i\xi Y_n(\alpha))\} - E^*\left\{|\bar{X}| < \epsilon, \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right\}\right|$$

$$\leq \frac{\alpha\xi\epsilon}{n} E^*\left\{|\bar{X}| < \epsilon, \sum_1^n (|X_k| + \epsilon)^{\alpha-1}\right\}.$$

Clearly $E^*\left\{|\bar{X}| < \epsilon, \sum_1^n (|X_k| + \epsilon)^{\alpha-1}\right\}$, approaches 0, as ϵ approaches 0, hence by (3)

$$E\{\exp(i\xi Y_n(\alpha))\} = \lim_{\epsilon\to 0} \frac{E^*\left\{|\bar{X}| < \epsilon, \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right\}}{\text{Prob}\{\,|\bar{X}| < \epsilon\}}. \tag{4}$$

Using the fact that

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin \epsilon\eta}{\eta}\exp(i\eta\bar{X})\,d\eta \begin{cases} = 1, & |\bar{X}| < \epsilon, \\ = \frac{1}{2}, & |\bar{X}| = \epsilon, \\ = 0, & |\bar{X}| > \epsilon, \end{cases}$$

we obtain easily

$$(5)\qquad \begin{aligned} &E^*\left\{ |\bar{X}| < \epsilon, \exp\left(\frac{i\xi}{n}\sum_1^n |X_k|^\alpha\right)\right\} \\ &\qquad = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin \epsilon\eta}{\eta} E\left\{ \exp\frac{i}{n}\sum_1^n (\xi\,|X_k|^\alpha + \eta X_k)\right\} d\eta. \end{aligned}$$

The justification of interchanging of the order of integration (from $-\infty$ to ∞) and the operation E can be made quite simply (see e.g. Kac and Steinhaus [3]). Notice now that

$$\begin{aligned} &E\left\{ \exp\frac{i}{n}\sum_1^n (\xi\,|X_k|^\alpha + \eta\, X_k)\right\} \\ &\qquad = \left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2}\right)\exp\frac{i}{n}(\xi\,|x|^\alpha + \eta x)\,dx\right]^n = \varphi_n(\xi, \eta) \end{aligned}$$

and that $\varphi_n(\xi, \eta)$ is absolutely integrable in $(-\infty, \infty)$ as a function of η. Thus, as $\epsilon \to 0$,

$$(6)\qquad \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin \epsilon\eta}{\eta}\varphi_n(\xi, \eta)\,d\eta \sim \frac{\epsilon}{\pi}\int_{-\infty}^{\infty}\varphi_n(\xi, \eta)\,d\eta.$$

Furthermore (as $\epsilon \to 0$)

$$(7)\qquad \text{Prob}\{\,|\bar{X}| < \epsilon\,\} \sim 2\epsilon\,\frac{\sqrt{n}}{\sqrt{2\pi}}$$

and combining this with (6), (5) and (4) we get

$$(8)\qquad E\{\exp(i\xi Y_n(\alpha))\} = \frac{1}{\sqrt{2\pi n}}\int_{-\infty}^{\infty}\varphi_n(\xi, \eta)\,d\eta.$$

This, of course, is equivalent to (2).

4. Density function of the mean deviation. If $\alpha = 1$ one can obtain an expression for the probability density $f_n(\beta)$ of $Y_n(\alpha)$. We note first that

$$\begin{aligned} &\int_{-\infty}^{\infty}\exp\left(-\frac{x^2}{2}\right)\exp\frac{i}{n}(\xi\,|x| + \eta x)\,dx \\ &\qquad = n\int_0^{\infty}\exp\left(-\frac{n^2x^2}{2}\right)\exp i(\xi + \eta)x\,dx \\ &\qquad\quad + n\int_0^{\infty}\exp\left(-\frac{n^2x^2}{2}\right)\exp i(\xi - \eta)x\,dx = n\{R(\xi + \eta) + R(\xi - \eta)\} \end{aligned}$$

where

$$R(u) = \int_0^\infty \exp\left(-\frac{n^2 x^2}{2}\right) \exp(iux)\, dx.$$

Using (2) (with $\alpha = 1$) we obtain

$$F_n(\xi) = \frac{n^n}{\sqrt{n}(\sqrt{2\pi})^{n+1}} \int_{-\infty}^\infty \left[\sum_{k=0}^n \binom{n}{k} R^k(\xi + \eta) R^{n-k}(\xi - \eta)\right] d\eta$$

Let us first look at the summands corresponding to $k = 0$ and $k = n$ We have

$$\int_{-\infty}^\infty R^n(\xi - \eta)\, d\eta = \int_{-\infty}^\infty R^n(\eta)\, d\eta = \int_{-\infty}^\infty R^n(\xi + \eta)\, d\eta.$$

Now, $R(\eta)$ is the Fourier transform of

$$\zeta(x) = \begin{cases} 0, & x < 0, \\ \exp\left(-\frac{n^2 x^2}{2}\right), & x > 0, \end{cases}$$

and hence $R^n(\eta)$ is the Fourier transform of the convolution

$$\underbrace{\zeta * \zeta * \cdots * \zeta}_{n} = \zeta^{(n)}(x).$$

It is easily seen (using integration by parts) that

$$R(\eta) = O\left(\frac{1}{|\eta|}\right)$$

for large $|\eta|$ and hence for $n \geq 2$, $R^n(\eta)$ is absolutely integrable in $(-\infty, \infty)$. It follows (classical inversion formula) that

$$\int_{-\infty}^\infty R^n(\eta)\, d\eta = 2\pi\zeta^{(n)}(0)$$

Since for $n \geq 2$, $\zeta^{(n)}(x)$ is continuous and $\zeta^{(n)}(x) = 0$ for $x < 0$ we have $\zeta^{(n)}(0) = 0$. Thus

$$F_n(\xi) = \frac{n^{n-\frac{1}{2}}}{(\sqrt{2\pi})^{n+1}} \sum_{k=1}^{n-1} \binom{n}{k} \int_{-\infty}^\infty R^k(\xi + \eta) R^{n-k}(\xi - \eta)\, d\eta.$$

It is fairly easy to check that

$$\int_{-\infty}^\infty R^k(\xi + \eta) R^{n-k}(\xi - \eta)\, d\eta = \pi \int_{-\infty}^\infty \exp(i\xi x)\zeta^{(k)}\left(\frac{x}{2}\right) \zeta^{(n-k)}\left(\frac{x}{2}\right) dx$$

so that

$$F_n(\xi) = \frac{\pi n^{n-\frac{1}{2}}}{(\sqrt{2\pi})^{n+1}} \int_{-\infty}^\infty \exp(i\xi x) \sum_{k=1}^{n-1} \binom{n}{k} \zeta^{(k)}\left(\frac{x}{2}\right) \zeta^{(n-k)}\left(\frac{x}{2}\right) dx$$

Finally,

$$f_n(\beta) = \frac{\pi n^{n-\frac{1}{2}}}{(\sqrt{2\pi})^{n+1}} \sum_{k=1}^{n-1} \binom{n}{k} \zeta^{(k)}\left(\frac{\beta}{2}\right) \zeta^{(n-k)}\left(\frac{\beta}{2}\right).$$

I have not been able, except for $n = 3$, to verify directly that this formula is identical with that of Goodwin.

REFERENCES

[1] H. J Goodwin, "On the distribution of the estimate of mean deviation obtained from samples from a normal population," *Biometrika*, Vol 33 (1945), pp 254–256

[2] J. F. Daly, "On the use of the sample range in an analogue of Student's t-test," *Annals of Math Stat.*, Vol 17 (1946), pp. 71–74.

[3] M Kac and H. Steinhaus, "Sur les fonctions indépendantes III," *Stud Math*, Vol 6 (1936), pp. 93–94.

NOTES

This section is devoted to brief research and expository articles and other short items.

A FUNCTIONAL EQUATION FOR WISHART'S DISTRIBUTION

BY G. RASCH

State Serum Institute and University of Copenhagen

1. Introduction. The sampling distribution of the moment matrix for observations from a multivariate normal distribution was given by Wishart in 1928 [1]. This proof involved rather advanced multidimensional geometry but since then two analytical proofs have been given: one by Wishart and Bartlett in cooperation with Ingham by the use of the characteristic function [2] and a second by Hsu by induction with regard to the dimension of the observations, [3],

In the following section is given a new derivation of the form of Wishart's distribution in which a fundamental property of the multivariate normal distribution is utilized, *viz.* the invariance of the distribution type against a linear transformation. In section 3 the same principle is used for evaluation of the constant and determination of the moment matrix in the multidimensional normal distribution.

2. Derivation of Wishart's distribution. Let[1]

$$\mathbf{x} = (x_1, \cdots, x_k), \tag{1}$$

denote a k-dimensional normal variate with the mean vector 0 and the distribution matrix

$$\mathbf{\Phi} = (\varphi_{ij}), \tag{2}$$

viz.

$$p\{\mathbf{x}\} = \frac{\sqrt{\Delta(\mathbf{\Phi})}}{(\sqrt{2\pi})^k} \cdot e^{-\frac{1}{2}\mathbf{x}\mathbf{\Phi}\mathbf{x}^*}. \tag{3}$$

$\mathbf{\Phi}$ is symmetrical and positive definite.

Now consider n observations of $\mathbf{x}$: $\mathbf{x}_1, \cdots, \mathbf{x}_n$, which are stochastically independent. Their joint distribution is

$$p\{\mathbf{x}_1, \cdots \mathbf{x}_n\} = \left(\frac{\sqrt{\Delta(\mathbf{\Phi})}}{(\sqrt{2\pi})^k}\right)^n \cdot e^{-\frac{1}{2}\Sigma\mathbf{x}_\nu\mathbf{\Phi}\mathbf{x}_\nu^*} \tag{4}$$

The estimation of $\mathbf{\Phi}$ is based upon the moment sums

$$m_{ij} = \Sigma x_{\nu i} x_{\nu j},$$

[1] Notations. Lower case latin and greek letters are scalars, boldface capital latin and greek letters denote matrices, and boldface lower case letters row vectors. * means transposition Δ (A) stands for the determinant of the square matrix **A**.

which form the symmetrical, positive definite matrix

$$\mathbf{M} = (m_{ij}) = \Sigma \mathbf{x}_\nu^* \mathbf{x}_\nu . \tag{5}$$

In order to derive the distribution of $\mathbf{M}$ the straightforward procedure seems to be to transform the distribution of the sample $(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ to a distribution of $\mathbf{M}$ and some other variables which then should be integrated away. As such, the transformation,

$$\mathbf{x}_\nu = \mathbf{u}_\nu \mathbf{M}^{\frac{1}{2}}, \qquad \mathbf{M}\Sigma \mathbf{u}_\nu^* \mathbf{u}_\nu = 1, \tag{6}$$

might serve. The matrix

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 \\ \cdots \\ \mathbf{u}_n \end{pmatrix} \tag{7}$$

contains nk elements linked together with $\frac{(k+1)k}{2}$ relations; $(\mathbf{U})$ symbolizes $\left(n - \frac{k+1}{2}\right) k$ of the elements taken as independent variables

For the purpose of introducing $\mathbf{M}$ in the exponential term in (4) we shall define the "double dot multiplication" of two matrices:

$$\mathbf{A} \cdot\cdot \mathbf{B} = (a_{ij}) \cdot\cdot (b_{ij}) = \sum_{(i)} \sum_{(j)} a_{ij} b_{ij}, \tag{8}$$

for which we notice the rule

$$\mathbf{A} \cdot\cdot (\mathbf{BCD}) = \mathbf{C} \cdot\cdot (\mathbf{B}^*\mathbf{A}\mathbf{D}^*). \tag{9}$$

As obviously

$$\mathbf{x}\Phi\mathbf{x}^* = \Sigma \varphi_{ij} x_i x_j = \Phi \cdot\cdot (\mathbf{x}^*\mathbf{x}),$$

we have

$$\Sigma \mathbf{x}_\nu \Phi \mathbf{x}_\nu^* = \Phi \cdot\cdot \mathbf{M}, \tag{10}$$

and accordingly

$$p\{\mathbf{M}, (\mathbf{U})\} = \left(\frac{\sqrt{\Delta(\Phi)}}{(\sqrt{2\pi})^k}\right)^n \cdot e^{-\frac{1}{2}\Phi \cdot\cdot \mathbf{M}} \left| \frac{\partial(\mathbf{x}_1, \cdots, \mathbf{x}_n)}{\partial(\mathbf{M}, (\mathbf{U}))} \right|, \tag{11}$$

where $\frac{\partial(\ )}{\partial(\ )}$ denotes the jacobian of the transformation. On integrating with respect to $(\mathbf{U})$ we obtain

$$p\{\mathbf{M}\} = (\sqrt{\Delta(\Phi)})^n \cdot e^{-\frac{1}{2}\Phi \cdot\cdot \mathbf{M}} \cdot \varphi(\mathbf{M}), \tag{12}$$

where $\varphi(\mathbf{M})$ is independent of Φ. From this it follows that $p\{\mathbf{x}_1, \cdots, \mathbf{x}_n \mid \mathbf{M}\}$ is independent of Φ, i.e. $\mathbf{M}$ is a sufficient statistic for Φ.

In order to determine the mathematical form of $\varphi(\mathbf{M})$ we shall apply an arbitrary linear transformation to the original variates:

$$\mathbf{x}_\nu = \mathbf{x}'_\nu \mathbf{A}. \tag{13}$$

The new variates x'_ν are obviously normally distributed about 0 with the distribution matrix

$$\Phi' = A\Phi A^*. \tag{14}$$

Therefore the distribution function of the new moment matrix, given by

$$M = A^*M'A, \tag{15}$$

is

$$p\{M'\} = (\sqrt{\Delta(\Phi')})^n \cdot e^{-\frac{1}{2}\Phi' \cdot\cdot M'} \varphi(M'). \tag{16}$$

On the other hand the transformation from M to M′ is a linear one, the jacobian of which therefore is a constant depending on A only:

$$\frac{\partial(M)}{\partial(M')} = \psi(A), \quad \text{say.} \tag{17}$$

Consequently,

$$p\{M'\} = \sqrt{\Delta(\Phi)} \cdot e^{-\frac{1}{2}\Phi \cdot\cdot M} \cdot \varphi(M) \cdot |\psi(A)|. \tag{18}$$

The two expressions for $p\{M'\}$ must be identical, and as

$$\Delta(\Phi') = \Delta(\Phi)\Delta^2(A), \tag{19}$$

and

$$\Phi' \cdot\cdot M' = (A\Phi A^*) \cdot\cdot M' = (A^*M'A) \cdot\cdot \Phi = M \cdot\cdot \Phi, \tag{20}$$

it follows that $\varphi(M)$ satisfies the functional equation

$$|\Delta(A)|^n \varphi(M') = \varphi(M) \cdot |\psi(A)|. \tag{21}$$

Now, since the transformation $M = (AB)^* M'(AB)$ may be carried out in two steps, $\psi(A)$ also satisfies a functional equation

$$\psi(AB) = \psi(A)\psi(B). \tag{22}$$

Furthermore, if A is a diagonal matrix it is easily seen that

$$\psi(A) = (\Delta(A))^{k+1}, \tag{23}$$

and this relation holds generally. In fact, considering the case where the normal form of A is a diagonal matrix:

$$A = TDT^{-1}, \text{ say,}$$

we get

$$\begin{aligned} \psi(A) &= \psi(T)\psi(D)\psi(T^{-1}) \\ &= (\Delta(D))^{k+1}\, \psi(TT^{-1}) \\ &= (\Delta(A))^{k+1}, \end{aligned}$$

and by analytical continuation this is seen to be true for any A.

Now, inserting this result in the functional equation (21) and taking for **A** the real symmetrical square root of **M** so that[2] $\mathbf{M}' = 1$, we readily obtain the solution

$$\varphi(\mathbf{M}) = (\Delta(\mathbf{M}^{\frac{1}{2}}))^{n-k-1}\, \varphi(1). \tag{24}$$

It follows that

$$p\{\mathbf{M}\} = \gamma_k(n)(\Delta(\Phi))^{n/2} \cdot e^{-\frac{1}{2}\Phi \cdot \cdot \mathbf{M}} \cdot (\Delta(\mathbf{M}))^{(n-k-1)/2}, \tag{25}$$

where $\gamma_k(n) = \varphi(1)$ is a constant which may be determined in various ways (cf. for instance Cramér [4]).

3. Other applications of the linear transformation. It may be noticed that the linear transformation also leads to simple derivations of two fundamental properties of the normal multivariate distribution itself, *viz.* determination of the constant and the relation between the moment matrix and the distribution matrix.

Let

$$p\{\mathbf{x}\} = \gamma(\Phi) \cdot e^{-\frac{1}{2}\mathbf{x}\Phi\mathbf{x}^*}, \tag{26}$$

and transform by

$$\mathbf{x} = \mathbf{x}'\mathbf{A}. \tag{27}$$

The new variable obviously has the distribution matrix (14) and the constant $\gamma(\Phi')$. But on the other hand direct transformation of (26) leads to

$$\begin{aligned} P\{\mathbf{x}'\} &= \gamma(\Phi) \cdot e^{-\frac{1}{2}\mathbf{x}\Phi\mathbf{x}^*} \cdot \left| \frac{\partial(\mathbf{x})}{\partial(\mathbf{x}')} \right| \\ &= \gamma(\Phi)\, |\, \Delta(\mathbf{A})\, |\, e^{-\frac{1}{2}\mathbf{x}\Phi\mathbf{x}^*}, \end{aligned}$$

and therefore we must have

$$\gamma(\Phi') = \gamma(\Phi)\, |\, \Delta(\mathbf{A})\, |\, .$$

For $\mathbf{A} = \Phi^{-\frac{1}{2}}$ we get $\Phi' = 1$ and consequently

$$\gamma(\Phi) = \sqrt{\Delta(\Phi)} \cdot \gamma(1),$$

where obviously

$$\gamma(1) = \frac{1}{(\sqrt{2\pi})^n}.$$

Considering

$$\mathbf{M}(\Phi) = \int \mathbf{x}^*\mathbf{x}p\{\mathbf{X}\}\, d\mathbf{x},$$

[2] Exists because **M** is positive definite. Let $\mathbf{M} = \mathbf{ODO}^*$ where **O** is orthogonal and **D** the diagonal form of **M**, then $\mathbf{M}^{\frac{1}{2}} = \mathbf{O}D^{\frac{1}{2}}\mathbf{O}^*$ is real and symmetrical

the same transformation gives

$$\mathbf{M}(\Phi) = \int \mathrm{A}^*\mathrm{x}^*\mathrm{x}\mathrm{A}p\{\mathrm{x}'\}\, d\mathrm{x}',$$

$$= \mathrm{A}^*\mathbf{M}(\Phi')\mathrm{A}$$

which for $A = \Phi^{-\frac{1}{2}}$ leaves us with

$$\mathbf{M}(\Phi) = (\Phi')^{-1}$$

because $\mathbf{M}(1) = 1$.

REFERENCES

[1] J. Wishart, "The generalised product moment distribution in samples from a normal multivariate population", *Biometrika*, Vol. 20A (1928), p. 32.
[2] J. Wishart and M S Bartlett, "The generalised product moment distribution in a normal system", *Proc. Cambr. Phil Soc.*, Vol. 29 (1933), p. 260.
[3] P. L. Hsu, "A new proof of the joint product moment distribution", *Proc. Cambr. Phil. Soc.*, Vol. 35 (1939), p. 336.
[4] H. Cramér, *Mathematical Methods in Statistics*, Princeton Univ. Press, 1946, pp. 392–93

THE DISTRIBUTION OF A DEFINITE QUADRATIC FORM

By Herbert Robbins

University of North Carolina

Let $X_1, \cdots, X_n$ be independent normal variates with zero means and unit variances, let $a_1, \cdots, a_n$ be positive constants and define

$$U_n = \frac{a_1}{2}X_1^2 + \cdots + \frac{a_n}{2}X_n^2, \tag{1}$$

$$F_n(x) = \Pr[U_n \le x], \qquad f_n(x) = F_n'(x). \tag{2}$$

Setting

$$a = (a_1 \cdots a_n)^{1/n} \tag{3}$$

and using the convolution formula we may show by induction that for $x > 0$,

$$f_n(x) = a^{-\frac{1}{2}n} x^{\frac{1}{2}n-1} \sum_{k=0}^{\infty} \frac{c_k(-x)^k}{\Gamma(\frac{1}{2}n + k)}, \tag{4}$$

$$F_n(x) = a^{-\frac{1}{2}n} x^{\frac{1}{2}n} \sum_{k=0}^{\infty} \frac{c_k(-x)^k}{\Gamma(\frac{1}{2}n + k + 1)}, \tag{5}$$

where for $k = 0, 1, \cdots$

$$c_k = \pi^{-\frac{1}{2}n} \sum_{i_1+\cdots+i_n=k} \frac{\Gamma(i_1 + \frac{1}{2})\cdots\Gamma(i_n + \frac{1}{2})}{i_1!\cdots i_n!\, a_1^{i_1}\cdots a_n^{i_n}} > 0. \tag{6}$$

In particular, if $a_1 = \cdots = a_n = 2$, then using the known distribution of χ^2 with n degrees of freedom we have

$$\bar{f}_n(x) = \frac{x^{\frac{1}{2}n-1}e^{-\frac{1}{2}x}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} = \frac{x^{\frac{1}{2}n-1}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}\sum_{k=0}^{\infty}\frac{(-x)^k}{2^k k!} = \frac{x^{\frac{1}{2}n-1}}{2^{\frac{1}{2}n}}\sum_{k=0}^{\infty}\frac{\bar{c}_k(-x)^k}{\Gamma(\frac{1}{2}n+k)},$$

so that

$$\bar{c}_k = \frac{\Gamma(\frac{1}{2}n+k)}{2^k k!\,\Gamma(\frac{1}{2}n)} = \frac{\pi^{-\frac{1}{2}n}}{2^k}\sum_{i_1+\cdots+i_n=k}\frac{\Gamma(i_1+\frac{1}{2})\cdots\Gamma(i_n+\frac{1}{2})}{i_1!\cdots i_n!},$$

and therefore

$$\sum_{i_1+\cdots+i_n=k}\frac{\Gamma(i_1+\frac{1}{2})\cdots\Gamma(i_n+\frac{1}{2})}{i_1!\cdots i_n!} = \frac{\pi^{\frac{1}{2}n}\Gamma(\frac{1}{2}n+k)}{k!\,\Gamma(\frac{1}{2}n)}. \tag{7}$$

Now in the general case let

$$\alpha = \min\{a_1, \cdots, a_n\}; \tag{8}$$

then from (6) and (7) we deduce that

$$\left|\frac{c_k(-x)^k}{\Gamma(\frac{1}{2}n+k)}\right| \le \frac{(x/\alpha)^k}{\Gamma(\frac{1}{2}n)k!}, \tag{9}$$

with a similar inequality for the general term of (5).

It is difficult to evaluate numerically the coefficient c_k by a direct application of the definition (6). We shall therefore give a method by which the c_k may be computed easily. We shall assume in what follows that the a_i are distinct.

Let $Y_1, \cdots, Y_n$ be another set of variates with the same joint distribution as the X_i and independent of the X_i, and set

$$V_{2n} = \frac{a_1}{2}X_1^2 + \cdots + \frac{a_n}{2}X_n^2 + \frac{a_1}{2}Y_1^2 + \cdots + \frac{a_n}{2}Y_n^2, \tag{10}$$

$$G_{2n}(x) = \Pr[V_{2n} \le x], \qquad g_{2n}(x) = G'_{2n}(x). \tag{11}$$

Then by the convolution formula,

$$g_{2n}(x) = \int_0^x f_n(x-y)f_n(y)\,dy = a^{-n}x^{n-1}\sum_{k=0}^{\infty}\left\{\sum_{i=0}^{k}c_i c_{k-i}\right\}\frac{(-x)^k}{\Gamma(k+n)}. \tag{12}$$

But we may show directly that, setting

$$q_i = \prod_{j\ne i}(a_j - a_i)^{-1} \qquad (i = 1, \cdots, n), \tag{13}$$

we have

$$g_{2n}(x) = (-1)^{n-1}\sum_{i=1}^{n}q_i a_i^{n-2}e^{-x/a_i} = (-1)^{n-1}\sum_{k=0}^{\infty}\left\{\sum_{i=1}^{n}q_i a_i^{n-k-2}\right\}\frac{(-x)^k}{k!}. \tag{14}$$

Equating coefficients in (12) and (14) we derive the fundamental formula

$$\sum_{i=0}^{k}c_i c_{k-i} = a^n\sum_{i=1}^{n}q_i a_i^{-(k+1)} \qquad (k = 0, 1, \cdots). \tag{15}$$

Thus, defining

$$2P_k = a^n \sum_{i=1}^{n} q_i a_i^{-(k+1)}, \tag{16}$$

we may write

$$\sum_{i=0}^{k} c_i c_{k-i} = 2P_k. \tag{17}$$

From (6) or (17) it follows that

$$c_0 = 1. \tag{18}$$

Thus we may solve (17) successively for the c_k in terms of the P_k: for $j = 0, 1, \cdots$

$$\begin{aligned} c_{2j} &= P_{2j} - \left\{ c_1 c_{2j-1} + c_2 c_{2j-2} + \cdots + c_{j-1} c_{j+1} + \frac{c_j^2}{2} \right\}, \\ c_{2j+1} &= P_{2j+1} - \{c_1 c_{2j} + c_2 c_{2j-1} + \cdots + c_j c_{j+1}\}. \end{aligned} \tag{19}$$

When the n constants $q_1, \cdots, q_n$ have been computed we may compute the P_k by (16) and then the c_k by (19) successively as far as desired. The inequality (9) gives an indication of the number of terms of the alternating series (4) or (5) which are necessary to secure a desired accuracy. A sharper result than (9) should certainly be possible when the a_i are well separated.

The foregoing method may be modified to cover cases in which some of the a_i are equal, the formulas (16) being replaced by the appropriate limits as the a_i approach equality.

We shall now derive an expansion of $f_n(x)$ and $F_n(x)$ in terms of χ^2 distributions. Let us set for $x > 0$,

$$f_n(x) = \sum_{k=0}^{\infty} (-1)^k d_k \cdot \frac{x^{\frac{1}{2}n+k-1} e^{-x/a}}{a^{\frac{1}{2}n+k} \Gamma(\frac{1}{2}n + k)}, \tag{20}$$

or, equivalently,

$$\begin{aligned} \frac{a}{2} f_n\left(\frac{a}{2} x\right) &= \sum_{k=0}^{\infty} (-1)^k d_k \frac{x^{\frac{1}{2}n+k-1} e^{-\frac{1}{2}x}}{2^{\frac{1}{2}n+k} \Gamma(\frac{1}{2}n + k)} \\ &= \frac{x^{\frac{1}{2}n-1} e^{-\frac{1}{2}x}}{2^{\frac{1}{2}n} \Gamma(\frac{1}{2}n)} \cdot \sum_{k=0}^{\infty} (-1)^k d_k \frac{(\frac{1}{2}x)^k \Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n + k)}, \end{aligned} \tag{21}$$

where the d_k are to be determined. It follows, after some reduction, that

$$g_{2n}(x) = a^{-n} x^{n-1} e^{-x/a} \sum_{k=0}^{\infty} \left\{ \sum_{i=0}^{k} d_i d_{k-i} \right\} \frac{(-x)^k}{a^k \Gamma(\frac{1}{2}n + k)}. \tag{22}$$

But we may write (14) in the form

$$g_{2n}(x) = a^{-n} x^{n-1} e^{-x/a} \sum_{k=0}^{\infty} \left\{ \sum_{i=1}^{n} a q_i a_i^{n-2-k} (a - a_i)^k \right\} \frac{(-x)^{k-n+1}}{a^{k-n+1} \Gamma(k + 1)}. \tag{23}$$

Equating coefficients in (22) and (23) and setting

$$2Q_k = a \sum_{i=1}^{n} q_i a_i^{-(k+1)} (a - a_i)^{n+k-1}, \tag{24}$$

we obtain the relations, $d_0 = 1$ and

$$\sum_{i=0}^{k} d_i d_{k-i} = 2Q_k \qquad (k = 0, 1, \cdots), \tag{25}$$

from which the d_k may be computed as in (19). Equation (20) or (21) then gives the expansion of $f_n(x)$ in a series of χ^2 frequency functions. The corresponding expansion of $F_n(x)$ is then

$$F_n(x) = \sum_{k=0}^{\infty} (-1)^k d_k \int_0^x \frac{t^{\frac{1}{2}n+k-1} e^{-t/a}}{a^{\frac{1}{2}n+k}\Gamma(\frac{1}{2}n + k)} dt, \tag{26}$$

or

$$F_n\left(\frac{a}{2} x\right) = \sum_{k=0}^{\infty} (-1)^k d_k \int_0^x \frac{t^{\frac{1}{2}n+k-1} e^{-t/2}}{2^{\frac{1}{2}n+k}\Gamma(\frac{1}{2}n + k)} dt. \tag{27}$$

By direct comparison of (4) and (20) we may establish the following relations among the c_k and d_k:

$$\begin{aligned} d_k &= (-1)^k \sum_{j=0}^{k} (-a)^j \binom{\frac{1}{2}n + k - 1}{k - j} c_j, \\ c_k &= a^{-k} \sum_{j=0}^{k} \binom{\frac{1}{2}n + k - 1}{k - j} d_j. \end{aligned} \tag{28}$$

These may be used if both the power series and the χ^2 series are desired.

From (6) we see directly that

$$c_1 = \frac{1}{2} \sum_{i=1}^{n} a_i^{-1}, \tag{29}$$

and from (28) it follows that

$$d_1 = \tfrac{1}{2} anb_1, \tag{30}$$

where we have set

$$b_1 = \left\{ \frac{1}{n} \sum_{i=1}^{n} a_i^{-1} \right\} - (a_1 \cdots a_n)^{-1/n}. \tag{31}$$

Since by a well known inequality $b_1 \geq 0$ it follows that $d_1 \geq 0$, with equality only if all the a_i are equal. If we denote by $h_n(x)$ the frequency function of $\frac{1}{2}a(X_1^2 + \cdots + X_n^2)$ then

$$\frac{a}{2} h_n\left(\frac{a}{2} x\right) = \frac{x^{\frac{1}{2}n-1} e^{-\frac{1}{2}x}}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}, \tag{32}$$

and hence (21) becomes

$$\frac{f_n(x)}{h_n(x)} = 1 - b_1 x + \cdots, \tag{33}$$

which is significant for x near 0.

EXACT LOWER MOMENTS OF ORDER STATISTICS IN SMALL SAMPLES FROM A NORMAL DISTRIBUTON

By Howard L. Jones

Illinois Bell Telephone Company

1. Summary. Exact means in samples of size ≤ 3, and exact second moments and product-moments in samples of size ≤ 4, are given in Table 1 in terms of π for order statistics selected from the normal distribution $N(0, 1)$. The derivation employs multiple integration and some general properties of the moments.

TABLE 1

Expected values of lower moments of order statistics, $x_i \geq x_{i+1}$, in samples of size n from the normal distribution $N(0, 1)$.

Moment	$n = 2$	$n = 3$	$n = 4$
$E[x_1]$	$1/\sqrt{\pi}$	$3/(2\sqrt{\pi})$	
$E[x_2]$		0	
$E[x_1^2]$	1	$1 + \sqrt{3}/(2\pi)$	$1 + \sqrt{3}/\pi$
$E[x_2^2]$		$1 - \sqrt{3}/\pi$	$1 - \sqrt{3}/\pi$
$E[x_1x_2]$	0	$\sqrt{3}/(2\pi)$	$\sqrt{3}/\pi$
$E[x_1x_3]$		$-\sqrt{3}/\pi$	$-(2\sqrt{3} - 3)/\pi$
$E[x_1x_4]$			$-3/\pi$
$E[x_2x_3]$			$(2\sqrt{3} - 3)/\pi$
σ_1^2	$1 - 1/\pi$	$1 - (9 - 2\sqrt{3})/(4\pi)$	
σ_2^2		$1 - \sqrt{3}/\pi$	
σ_{12}	$1/\pi$	$\sqrt{3}/(2\pi)$	
σ_{13}		$(9 - 4\sqrt{3})/(4\pi)$	

2. Introduction. The usefulness of the lower moments of order statistics for determining the moments of the range and for other purposes is well established. In small samples, however, computation of the moments by quadrature is laborious [1]. The values shown in Table 1 should therefore be helpful in problems requiring the use of these moments for samples of size ≤ 4, since the constant π has been evaluated to several hundred decimal places. Some of the methods used to obtain these results may also be useful in approximating or verifying the moments in larger samples.

3. Multiple integration. Let n random selections from the normal distribution $N(0, 1)$ be arranged in order of size so that

$$x_1 \geq x_2 \geq \cdots \geq x_n .$$

For samples of size 2, the means and product-moment are easily obtained from the general formula

$$E[x_i^k x_j^h] = n! \int_{-\infty}^{\infty} \int_{x_n}^{\infty} \int_{x_{n-1}}^{\infty} \cdots \int_{x_2}^{\infty} x_i^k x_j^h$$
$$f(x_1)f(x_2) \cdots f(x_n)\, dx_1\, dx_2 \cdots dx_n$$

where

$$f(x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_i^2},$$

$E\,[x_i^k]$ being the special case where $h = 0$. Multiple integration can also be used to find any product-moment, $E\,[x_i x_{j \neq i}]$, for samples of size 3, the order of integration being changed at any stage where necessary.

For the means in samples of size 3 and the product-moments in samples of size 4, the integrals reduce to double integrals which can be evaluated from the equation

$$\int_{-\infty}^{\infty} \int_{t_2}^{\infty} e^{-(a^2t_1^2+b^2t_2^2)}\, dt_1\, dt_2 = \frac{\pi}{2ab}.$$

This equation follows from the fact that

$$\int_{-\infty}^{\infty} \int_{bt_2/a}^{\infty} \frac{ab}{\pi} e^{-(a^2t_1^2+b^2t_2^2)}\, dt_1\, dt_2$$

is equivalent to

$$\int_0^1 \int_{p_2}^1 dp_1\, dp_2 ,$$

while the function

$$\phi(t_2) = e^{-b^2t_2^2} \int_{t_2}^{bt_2/a} e^{-a^2t_1^2}\, dt_1$$

has the symmetrical property that $\phi\,(t_2) = -\phi(-t_2)$, whence

$$\int_{-\infty}^{\infty} \phi(t_2)\, dt_2 = 0.$$

4. Some properties of the moments. The most obvious property of the moments of order statistics in samples from the normal distribution $N(0, 1)$ is their symmetry; thus:

$$E\,[x_i] = -E\,[x_{n-i+1}],$$
$$E\,[x_i^2] = E\,[x_{n-i+1}^2],$$
$$E\,[x_i x_j] = E\,[x_{n-i+1}x_{n-j+1}].$$

When sample values from any parent distribution are numbered in order of random selection, x_i and $x_{j \neq i}$ are statistically independent of each other, and the expected value of a product $x_i^k x_j^h$ is the product of the expected values of x_i^k and x_j^h. Numbering in order of size has the effect of increasing some expected values and decreasing others, leaving the sum of expected values of a given type unchanged, so that in general,

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (E[x_i^k x_j^h]) = \binom{n}{2} E[x_0^k]E[x_0^h]$$

where x_0 is a random selection. In particular, this equation holds for the special cases ($k = 1, h = 1$), ($k = 1, h = 0$), and ($k = 2, h = 0$), so that in samples from the normal distribution $N(0, 1)$,

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (E[x_i x_j]) = \tfrac{1}{2}n(n - 1)(E[x_0])^2 = 0,$$

$$\sum_{i=1}^{n} (E[x_i]) = nE[x_0] = 0,$$

$$\sum_{i=1}^{n} (E[x_i^2]) = nE[x_0^2] = n.$$

The foregoing relationships lead immediately to the evaluation of $E[x_1 x_2]$ and $E[x_1^2]$ in samples of size 2. (The generalization of these relationships was suggested by Professor John H. Smith, whose unpublished manuscript on sampling from a rectangular distribution has also been instructive.)

In samples from a normal distribution, the covariance of every order statistic with the sample mean is the same as the variance of the sample mean. This implies that the variance of the sample mean $\leq$ the variance of any order statistic, the ratio of one standard deviation to the other being equal to the coefficient of correlation between the sample mean and the order statistic. To derive these properties, consider the linear function

$$m = w_1 X_1 + w_2 X_2 + \cdots + w_n X_n$$

of the order statistics $X_1, X_2, \cdots, X_n$ in a sample selected from the normal distribution $N(\mu, \sigma)$ with unknown μ and σ. Let

$$x_i = (X_i - \mu)/\sigma, \qquad i = 1, 2, \cdots, n.$$

The conditions $w_1 + w_2 + \cdots + w_n = 1$ and $w_{n-i+1} = w_i$ are sufficient to make m an unbiased estimate of μ with variance $\sigma^2 E[(w_1 x_1 + w_2 x_2 + \cdots + w_n x_n)^2]$. The w's that make this variance minimum must satisfy the equations obtained by replacing w_i with w_{n-i+1}, for $i > \frac{1}{2}(n + 1)$, in the expression

$$E[(w_1 x_1 + w_2 x_2 + \cdots + w_n x_n)^2] + \lambda(w_1 + w_2 + \cdots + w_n - 1)$$

and then setting the partial derivative with respect to each w equal to zero. This leads to

$$\sum_{j=1}^{n} w_j E[x_i x_j] + \sum_{j=1}^{n} w_j E[x_{n-i+1} x_j] + \lambda = 0, \qquad 1 \leq i \leq n,$$

where the summations include the terms $E[x_i^2]$ and $E[x_{n-i+1}^2]$, respectively. But it is known [2] that the sample mean is the regular unbiased estimate of μ with minimum variance. Setting each w equal to $1/n$ and combining equivalent terms yields

$$\sum_{j=1}^{n} E[x_i x_j] + \tfrac{1}{2}n\lambda = 0, \qquad i = 1, 2, \cdots, n.$$

Summing from $i = 1$ to $i = n$, and employing the relationships discussed in the preceding paragraph, we obtain

$$n + \tfrac{1}{2}n^2\lambda = 0,$$

whence

$$\lambda = -2/n,$$

and

$$\sum_{j=1}^{n} E[x_i x_j] = 1, \qquad i = 1, 2, \cdots, n,$$

where the summation includes the term $E[x_i^2]$. This equation leads to the properties mentioned at the beginning of this paragraph. The same equation can be used to evaluate $E[x_1^2]$ and $E[x_2^2]$ in samples of size 3 or 4 from the distribution $N(0, 1)$, after the product-moments have been found.

REFERENCES

[1] C. Hastings, Jr., F. Mosteller, J. W. Tukey, and C. P. Winsor, "Low moments for small samples; a comparative study of order statistics," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 413–426.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946, p. 483.

NOTE ON AN ASYMPTOTIC EXPANSION OF THE nTH DIFFERENCE OF ZERO

By L. C. Hsu

National Tsing-Hua University, Peiping, China

This note gives an asymptotic expansion of the nth difference of zero. It is known that the Stirling number $S_{n,s}$ of the second kind is defined by

$$(1) \qquad n!\,S_{n,s} = \Delta^n 0^s = \sum_{x=0}^{n} (-1)^{n-x} \binom{n}{x} x^s.$$

We shall first show that the Stirling number $S_{n,n+k}$ can be expanded in the form

$$(2) \qquad S_{n,n+k} = \frac{n^{2k}}{2^k \cdot k!}\left[1 + \frac{f_1(k)}{n} + \frac{f_2(k)}{n^2} + \cdots + \frac{f_t(k)}{n^t} + O(n^{-t-1})\right], \quad (t < k)$$

where $f_1, f_2, \cdots, f_l$ are polynomials in k and whose coefficients can be found by means of the following lemmas.

The first lemma is due to B. F. Kimball, [1, (5 3)]

LEMMA 1. (Kimball) *Let q be a real number such that $n + q > 0$, and let $f(x) = x^{n+q}$. Then we can write $\Delta^n f(x)$ in the form*

$$\Delta^n f(x) = f^{[n]}(x + \tfrac{1}{2}n)\left[1 + \sum_{m=1}^{\infty} \binom{q}{2m}\cdot\left(\frac{n}{2x+n}\right)^{2m} W(m, n)\right], \tag{3}$$

where the value of $W(m, n)$ is given by

$$W(m, n) = B_{2m}^{-n}(-\tfrac{1}{2}n)/(\tfrac{1}{2}n)^{2m}, \tag{4}$$

$B_\nu^{-n}(x)$ being a so-called Bernoulli polynomial of negative order which was first defined by Norlund [2].

LEMMA 2. *Let the sum of all $\binom{n}{k}$ products of k different numbers taken from the set $(1, 2, \cdots, n)$ be denoted by $S_k(n)$. Then we can express it in the form*

$$S_k(n) = \sum_{\rho=1}^{k} (-1)^{k+\rho}\lambda_\rho(k)\binom{n+\rho}{k+\rho}, \tag{5}$$

where the coefficients $\lambda_1(k), \lambda_2(k), \cdots$ satisfy the recurrence relation

$$(k + \rho)\lambda_{\rho-1}(k) + \rho\cdot\lambda_\rho(k) = \lambda_\rho(k + 1) \tag{6}$$

with $\lambda_0 \equiv 0$, $\lambda_1 \equiv 1$ and $\lambda_{k+1}(k) = 0$.

PROOF. Clearly, among all $\binom{n}{k+1}$ products of $(k + 1)$ numbers out of $(1, 2, \cdots, n)$, there are exactly $\binom{n-1}{k}$ products containing the greatest factor n. The sum of these products is therefore $n\cdot S_k(n - 1)$. Repeating this reasoning, we get

$$S_{k+1}(n) = n\cdot S_k(n-1) + (n-1)\cdot S_k(n-2) + \cdots + (k+1)\cdot S_k(k). \tag{7}$$

Evidently, (5) is true for $k = 1$. Suppose now that it is true for $k = k$. Then the right-hand side of (7) can be written as

$$\begin{aligned}\sum_{\mu=0}^{n-k-1} (n - \mu) \sum_{\rho=1}^{k} (-1)^{k+\rho}\lambda_\rho(k)&\binom{n+\rho-\mu-1}{k+\rho} \\ &= \sum_{\rho=1}^{k}(-1)^{k+\rho}\lambda_\rho(k)\left[(k+\rho+1)\binom{n+\rho+1}{k+\rho+2} - \rho\binom{n+\rho}{k+\rho+1}\right] \\ &= \sum_{\rho=1}^{k+1}(-1)^{k+\rho+1}[(k+\rho)\lambda_{\rho-1}(k) + \rho\cdot\lambda_\rho(k)]\binom{n+\rho}{k+\rho+1}.\end{aligned}$$

The lemma thus follows by induction on k.

The number $S_k(n)$ may be called a Stirling number of the first kind. By the lemma just proved, it is easy to find

$$S_2(n) = 3\binom{n+2}{4} - \binom{n+1}{3}$$

$$S_3(n) = 15\binom{n+3}{6} - 10\binom{n+2}{5} + \binom{n+1}{4}$$

$$(8)\quad S_4(n) = 105\binom{n+4}{8} - 105\binom{n+3}{7} + 25\binom{n+2}{6} - \binom{n+1}{5}$$

$$S_5(n) = 945\binom{n+5}{10} - 1260\binom{n+4}{9} + 490\binom{n+3}{8} - 56\binom{n+2}{7} + \binom{n+1}{6}.$$

We shall see that in order to compute the coefficients of $f_1(k), f_2(k), \cdots$, it is sufficient to compute the values of $W(m, n), \lambda_1(m), \lambda_2(m), \cdots, (m = 1, 2, \cdots, t)$.

Let $f(x) = x^{n+k}$. Then by lemma 1, we have

$$n!\,S_{n,n+k} = \left[\frac{d^n}{dx^n} f(x + \tfrac{1}{2}n)\right]_{x=0} \cdot \left[1 + \sum_{m=1}^{\infty} \binom{k}{2m} W(m, n)\right].$$

From the definition of $S_k(n)$ it is easily seen that

$$(n + k)(n + k - 1) \cdots (n + 1) = n^n + n^{n-1}S_1(k) + \cdots + nS_{k-1}(k) + S_k(k).$$

Hence we may write

$$\frac{1}{n!}\left[\frac{d^n}{dx^n} f(x + \tfrac{1}{2}n)\right]_{x=0} = \frac{n^{2k}}{2^k \cdot k!}\left(1 + \frac{S_1(k)}{n} + \frac{S_2(k)}{n^2} + \cdots + \frac{S_k(k)}{n^k}\right).$$

It is clear from Kimball's paper [1] that

$$\sum_{m=1}^{\infty} \binom{k}{2m} W(m, n) = \sum_{n=1}^{t} \binom{k}{2m} W(m, n) + O(n^{-t-1}).$$

Substituting, we obtain

$$(9)\quad \begin{aligned} S_{n,n+k} &= \frac{n^{2k}}{2^k \cdot k!}\left[1 + \sum_{m=1}^{t} \binom{k}{2m} W(m, n) + O(n^{-t-1})\right] \\ &\qquad \cdot \left[1 + \sum_{m=1}^{t} \frac{S_m(k)}{n^m} + O(n^{-t-1})\right] \\ &= \frac{n^{2k}}{2^k \cdot k!}\left[1 + \sum_{m=1}^{t} \binom{k}{2m} W(m, n) + O(n^{-t-1})\right] \\ &\qquad \cdot \left[1 + \sum_{m=1}^{t} \sum_{\rho=1}^{m} \binom{k+\rho}{m+\rho} \frac{\lambda_\rho(m)}{(-1)^\rho(-n)^m} + O(n^{-t-1})\right]. \end{aligned}$$

The last expression shows that the asymptotic expansion (2) can be obtained by computing the numbers $\lambda_\rho(m)$, $W(m, n)$ with $1 \le \rho \le m \le t$. For example, consider the case $t = 3$ and notice that [1, (2 13)]

$$B_2^{-n}\left(-\frac{n}{2}\right) = \frac{n}{12}, \qquad B_4^{-n}\left(-\frac{n}{2}\right) = \frac{n^2}{48} - \frac{n}{120}, \qquad B_6^{-n}\left(-\frac{n}{2}\right) = \frac{5n^3}{576} + O(n^2),$$

and that $\lambda_1 \equiv 1$, $\lambda_2(2) = 3$, $\lambda_2(3) = 10$, $\lambda_3(3) = 15$. Then by a straightforward calculation of the right-hand side of (9) and by comparison with (2), we find

$$f_1(k) = \frac{1}{3}(2k^2 + k)$$

$$f_2(k) = \frac{1}{18}(4k^4 - k^2 - 3k) \tag{10}$$

$$f_3(k) = \frac{1}{810}(40k^6 - 60k^5 - 2k^4 - 63k^3 + 133k^2 - 48k).$$

Finally, combining (2) with the well-known Stirling's formula [3]

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\left[1 + \frac{1}{12n} + \frac{1}{288n^2} - \frac{139}{51840n^3} + O(n^{-4})\right], \tag{11}$$

and noting (1), we obtain

$$\Delta^n 0^{n+k} = \frac{\sqrt{2\pi n}}{k!}\left(\frac{n^2}{2}\right)^k\left(\frac{n}{e}\right)^n\left[1 + \frac{g_1(k)}{n} + \frac{g_2(k)}{n^2} + \frac{g_3(k)}{n^3} + O(n^{-4})\right] \tag{12.1}$$

where $g_1(k)$, $g_2(k)$, $g_3(k)$ are polynomials in k, viz.

$$g_1(k) = \frac{1}{12}(8k^2 + 4k + 1).$$

$$g_2(k) = \frac{1}{288}(64k^4 - 40k + 1) \tag{12.2}$$

$$g_3(k) = \frac{1}{51840} 2560k^6 - 3840k^5 + 832k^4 - 4032k^3 + 8392k^2 - 3732k - 139).$$

The asymptotic formula of $\Delta^n 0^{n+k}$ just derived is much better than a result previously obtained [4]. Moreover, it may be noted that the asymptotic expansion of $S_{n,n+k}$ may be made as sharp as desired, since in fact, for any prescribed $t > 1$, $\lambda_\rho(m)$ and $B_{2m}^{-n}(-\frac{1}{2}n)$, $(1 \le m \le t)$, may be easily computed by (6) and Kimball's [1, (2 12)] respectively

REFERENCES

[1] B F Kimball, "The application of Bernoulli polynomials of negative order to differencing." *Amer Math Jour.*, Vol 55 (1933), pp 399–416

[2] N E Nörlund *Differenzenrechnung*, Julius Springer, 1924, p 138.

[3] E T WHITTAKER AND G. N WATSON, *Modern Analysis*, Cambridge Univ Press, p 253.
[4] L. C HSU, "Some combinatorial formulas with applications to mathematical expectations and to differences of zero," *Annals of Math. Stat.*, Vol. 15 (1944)

AN INEQUALITY FOR KURTOSIS

BY LOUIS GUTTMAN

Cornell University

1. Summary. It is well known that, if the fourth moment about the mean of a frequency distribution equals the square of the variance, then the frequencies are piled up at exactly two points, namely, the two points that are one standard deviation away from the mean In this paper is developed a general inequality which describes the piling up of frequency around these two points for the case where the fourth moment exceeds the square of the variance. In a sense, it is shown how "U-shaped" a distribution must be according to its second and fourth moments.

2. An inequality. Let x be a random variable whose distribution has the following moments:

$$\mu = E(x); \sigma^2 = E(x - \mu)^2; (\alpha^2 + 1)\sigma^4 = E(x - \mu)^4. \tag{1}$$

α^2 is non-negative for any distribution, and its positive square root will be denoted by α. Let

$$t = (x - \mu)/\sigma \tag{2}$$

It will be shown that, if λ is an arbitrary positive number, then

$$\text{Prob } \{1 - \lambda\alpha \leqq t^2 \leqq 1 + \lambda\alpha\} > 1 - \lambda^{-2}. \tag{3}$$

If λ is chosen so as to make the left member in the braces positive, then t^2 is bounded away from zero, and (3) becomes:

$$\text{Prob } \{\sqrt{1 - \lambda\alpha} \leqq |t| \leqq \sqrt{1 + \lambda\alpha}\} > 1 - \lambda^{-2}, \qquad (\lambda\alpha < 1). \tag{4}$$

For example, if $\alpha = .5$ and $\lambda = \sqrt{2}$, then (4) shows that the probability is greater than .50 that t is either between .54 and 1 30, or between -1 30 and $-.54$. If $\alpha =$ 2 and $\lambda = 3$, then (4) shows that the probability is greater than .88 that t is either between 63 and 1.27, or between -1.27 and $-.63$. In general, the smaller α is, the greater the probability that t is in a small interval around $+1$ or -1. In particular, if $\alpha = 0$, then λ may be taken arbitrarily large, so that (4) shows that the probability is unity that $t = \pm 1$; this is the well known case referred to above.

3. Derivation. Inequality (3) is a special case of a slightly more general inequality which follows very simply from that of Tchebychef. Consider the function $t^2 - 1 + c$, where c is an arbitrary real number. By using (1) and (2), it is seen that

$$E(t^2 - 1 + c)^2 = \alpha^2 + c^2. \tag{5}$$

Then, according to Tchebychef's inequality, if λ is an arbitrary positive number,

$$\text{Prob}\ \{(t^2 - 1 + c)^2 \leqq \lambda^2(\alpha^2 + c^2)\} > 1 - \lambda^{-2}, \tag{6}$$

or,

$$\text{Prob}\ \{1 - c - \lambda\sqrt{\alpha^2 + c^2} \leqq t^2 \leqq 1 - c + \lambda\sqrt{\alpha^2 + c^2}\} > 1 - \lambda^{-2}. \tag{7}$$

This is the general inequality that was to be shown.

Inequality (3) is obtained by setting $c = 0$ in (7).

Another special case is obtained by determining c so as to maximize the left member in the braces of (7) By differentiation, the maximizing value is found to be $c = -\alpha/\sqrt{\lambda^2 - 1}$, for which (7) becomes:

$$\text{Prob}\ \{1 - \alpha\beta \leqq t^2 \leqq 1 + \alpha(\beta^2 + 2)/\beta\} > 1 - 1/(\beta^2 + 1), \tag{8}$$

where β is used instead of the notation $\sqrt{\lambda^2 - 1}$, and denotes any positive number. For the same probability on the right, (8) has the advantage over (3) of having $1 - \alpha\beta$ greater than $1 - \lambda\alpha$, so that the former may be positive even though the latter is negative. Inequality (8) starts the positive interval for t as close to $+1$ as possible. On the hand, (3) provides the minimum size interval for t^2 from among all values of c that make the left member in the braces of (7) positive.

If it is desired to have the positive interval for t *end* as close to $+1$ as possible, then the right member in the braces of (7) is to be minimized. By differentiation, the minimizing value is found to be $c = \alpha/\sqrt{\lambda^2 - 1}$, and the minimum inequality is:

$$\text{Prob}\ \{1 - \alpha(\beta^2 + 2)/\beta \leqq t^2 \leqq 1 + \alpha\beta\} > 1 - 1/(\beta^2 + 1). \tag{9}$$

4. Distribution Around μ. If the left member in the braces of (7) is negative, then instead of giving information about the piling up of probability of t around $+1$ and -1, (7) provides a statement about the probability in an interval around μ. Alternatively, this may be regarded as a confidence interval for μ. The minimum interval is given by (9); actually, it holds regardless of the value of the left member in the braces, another way of stating it is:

$$\text{Prob}\ \{-\sqrt{1 + \alpha\beta} \leqq t \leqq \sqrt{1 + \alpha\beta}\} > 1 - 1/(\beta^2 + 1). \tag{10}$$

TABLE FOR ESTIMATING THE GOODNESS OF FIT OF EMPIRICAL DISTRIBUTIONS

By N. Smirnov

1. Editorial Note. The table presented on pp 280-281 was originally published in [1]. It gives values of

$$L(z) = 1 - 2\sum_{\nu=1}^{\infty} (-1)^{\nu-1} e^{-\nu^2 z^2} = (2\pi)^{\frac{1}{2}} z^{-1} \sum_{\nu=1}^{\infty} e^{-(2\nu-1)^2\pi^2/8z^2},$$

which is also derived in [2].

Let $(X_1, \cdots, X_n)$ be a sample of independent variables with the same continuous cumulative distribution function $F(x)$, and let $N(z)$ be the number of X_k which are $\leq z$ By empirical distribution is meant the step-function $F_n^*(z) = N(z)/n$. The maximum D_n of the difference $|F_n^*(z) - F(z)|$ is a random variable and $L(z)$ is the limiting cumulative distribution function of $n^{1/2}D_n$. If $D_{m,n}$ is the maximum of the difference $|F_m^*(z) - F_n^{**}(z)|$ between the empirical distributions of two independent samples of sizes m and n, respectively, then $L(z)$ is also the limiting cumulative distribution function of $(mn/(m+n))^{1/2}D_{m,n}$.

REFERENCES

[1] N. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," *Bulletin Mathématique de l'Université de Moscou*, Vol. 2 (1939), fasc. 2.

[2] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distributions," *Annals of Math. Stat.*, Vol. 19 (1948), pp 177–189.

TABLE of $L(z)$

z	$L(z)$
.28	.000 001
.29	.000 004
.30	.000 009
.31	.000 021
.32	.000 046
.33	000 091
.34	.000 171
.35	.000 303
.36	000 511
.37	.000 826
.38	.001 285
39	.001 929
.40	.002 808
.41	.003 972
.42	.005 476
.43	.007 377
.44	.009 730
.45	.012 590
.46	.016 005
.47	.020 022
.48	.024 682
.49	.030 017
.50	.036 055
.51	.042 814
.52	.050 306
.53	058 534
.54	.067 497
.55	.077 183
56	.087 577
.57	.098 656
.58	.110 395
59	.122 760
.60	.135 718
.61	.149 229
.62	.163 225
63	.177 753
.64	192 677
.65	.207 987
66	.223 637
67	.239 582
.68	255 780

TABLE of $L(z)$—Continued

z	$L(z)$
.69	.272 189
70	.288 765
.71	.305 471
.72	.322 265
.73	.339 113
.74	.355 981
.75	.372 833
76	.389 640
.77	.406 372
.78	.423 002
.79	.439 505
.80	.455 857
.81	472 041
.82	.488 030
83	.503 808
.84	.519 366
.85	.534 682
.86	.549 744
.87	.564 546
.88	.579 070
.89	593 316
.90	.607 270
.91	.620 928
.92	.634 286
93	.647 338
.94	.660 082
.95	.672 516
.96	.684 636
.97	.696 444
.98	.707 940
.99	.719 126
1.00	.730 000
1.01	740 566
1.02	.750 826
1.03	.760 780
1.04	.770 434
1.05	.779 794
1.06	.788 860
1.07	.797 636
1.08	.806 128

TABLE of $L(z)$—Continued

z	$L(z)$
1.09	.814 342
1.10	.822 282
1.11	.829 950
1.12	.837 356
1.13	.844 502
1.14	.851 394
1 15	.858 038
1.16	.864 442
1.17	.870 612
1.18	.876 548
1.19	.882 258
1.20	.887 750
1.21	.893 030
1.22	.898 104
1.23	.902 972
1.24	.907 648
1.25	912 132
1.26	.916 432
1.27	.920 556
1.28	.924 505
1.29	.928 288
1.30	.931 908
1.31	.935 370
1.32	.938 682
1.33	.941 848
1.34	.944 872
1.35	.947 756
1.36	.950 512
1.37	.953 142
1.38	.955 650
1.39	.958 040
1.40	.960 318
1.41	.962 486
1.42	.964 552
1.43	966 516
1.44	.968 382
1.45	.970 158
1 46	.971 846
1.47	.973 448
1.48	.974 970

TABLE of $L(z)$—*Continued*

z	$L(z)$
1.49	.976 412
1.50	.977 782
1.51	.979 080
1.52	.980 310
1.53	.981 476
1.54	.982 578
1.55	.983 622
1.56	.984 610
1.57	.985 544
1 58	.986 426
1.59	.987 260
1.60	.988 048
1.61	.988 791
1.62	.989 492
1.63	.990 154
1.64	.990 777
1.65	991 364
1.66	991 917
1 67	.992 438
1.68	.992 928
1.69	.993 389
1.70	.993 823
1.71	994 230
1.72	.994 612
1.73	.994 972
1.74	.995 309
1.75	.995 625
1.76	995 922
1.77	.996 200
1 78	.996 460
1.79	.996 704
1 80	.996 932
1.81	997 146
1.82	.997 346
1.83	997 533
1.84	.997 707
1.85	.997 870
1.86	.998 023
1.87	.998 145
1.88	.998 297

TABLE of $L(z)$—*Continued*

z	$L(z)$
1.89	.998 421
1.90	.998 536
1.91	.998 644
1.92	.998 744
1.93	.998 837
1.94	.998 924
1.95	.999 004
1 96	.999 079
1.97	.999 149
1.98	.999 213
1.99	.999 273
2.00	.999 329
2.01	.999 380
2.02	.999 428
2.03	.999 474
2.04	.999 516
2.05	.999 552
2.06	.999 588
2.07	.999 620
2.08	.999 650
2.09	.999 680
2.10	.999 705
2.11	.999 728
2.12	.999 750
2.13	999 770
2.14	.999 790
2.15	999 806
2.16	.999 822
2.17	999 838
2.18	999 852
2.19	.999 864
2.20	.999 874
2.21	.999 886
2.22	.999 896
2.23	.999 904
2.24	.999 912
2.25	.999 920
2.26	.999 926
2.27	.999 934
2.28	.999 940

TABLE of $L(z)$—*Concluded*

z	$L(z)$
2.29	.999 944
2.30	.999 949
2.31	.999 954
2.32	.999 958
2.33	.999 962
2.34	.999 965
2.35	.999 968
2.36	.999 970
2.37	.999 973
2.38	.999 976
2.39	.999 978
2.40	.999 980
2.41	.999 982
2.42	.999 984
2.43	.999 986
2.44	.999 987
2.45	.999 988
2.46	.999 989
2.47	.999 990
2 48	.999 991
2.49	.999 992
2.50	.999 9925
2 55	.999 9956
2 60	.999 9974
2.65	.999 9984
2.70	.999 9990
2.75	.999 9994
2 80	.999 9997
2.85	.999 99982
2.90	.999 99990
2.95	.999 99994
3.00	999 99997

BOOK REVIEW

Fundamentals of Statistics *Truman Lee Kelley.* Harvard University Press, 1947; pp. xvi, 755. $10.00.

REVIEWED BY A. M. MOOD

Iowa State College

First, a brief look at the contents: introductory matter, broad classifications of types of data, quantitative and qualitative aspects of data, construction of tables, charts, and graphs—200 pages; location and scale parameters, and moments—75 pages; normal distribution—30 pages; exact sampling distributions based on normal theory—5 pages; binomial distribution, goodness of fit tests, contingency tables, normal approximation to the distribution of the variance ratio, properties of Chi-square—20 pages; correlation and regression—150 pages.

These first 480 pages constitute the essential part of the book and the part that will be commented on here. But there are 270 more pages, the content of which we shall merely note without comment. There is a chapter of 90 pages entitled "Sundry Statistical Issues and Procedures" which discusses fifteen issues such as periodicity, time series, curve fitting, variance error of a coefficient corrected for attenuation, machine extraction of square roots, and sequential analysis. There follows a chapter of 40 pages devoted to no less than twenty-three topics in mathematics, topics such as: matrices and determinants, the square root transformation, expanding a table, spaces of three or more dimensions, and Fourier series. The remaining 140 pages contain numerical tables, references, various indexes, and a test designed to measure the adequacy of students' mathematical preparation.

This then is another book which deals with the descriptive aspects of statistics. Despite its title, it omits discussion of distribution theory, sampling theory, the theory of estimation, tests of hypotheses, or the theory of probability. The phrase "confidence interval" appears not once, I believe, in the entire 750 pages. The discussion of Student's distribution is brief enough to be quoted in its entirety (page 284): "The t-distribution, shown through the courtesy of Dr. Philip J. Rulon, in Chart VIII II, is appropriate for interpreting the significance of means, differences of means, and of regression coefficients, for small samples—say N less than 15. It is the distribution of these statistics computed from small samples drawn from a parent normal distribution."

Thus the author denies any value to the developments in the fundamentals of statistics during the past twenty-five or thirty years. He does this not merely by implication but in so many words, referring to modern statistical inference, he writes (page 13): "A still greater weakness is that it is essentially a deductive procedure and relatively sterile in suggesting new courses—in inspiring creative inferences. It is fundamentally a method of proof and not one of invention."

He is therefore fully aware of his extreme position, and takes great pains to justify it. His thesis is that the main purpose of statistics is to suggest new hypotheses to the scientist. In developing this thesis he writes (page 15): "The physicist observes seemingly irregular changes in x as y changes. He repeats his experiment, controlling more and more of the conditions, and repeats again and again, and, if successful, he reaches a law at the end of his work. He has been using statistics." But his discussion avoids certain relevant questions. Why does the physicist repeat the experiment? Why did he perform it in the first place? Did he suspect before he collected any data that x and y might be related?

At any rate, the opinion of most present-day statisticians is that the primary role of statistics in scientific research is statistical inference. This opinion is certainly well-founded in my own experience. Here at Iowa State College the Statistical Laboratory is intimately implicated in the research programs of all departments—physical, biological, and social. These scientists perform their experiments with a specific purpose in mind—usually the estimation of some parameters, sometimes the testing of a hypothesis. They never seem to seek in a collection of data some new hypothesis by artful selection between the mean, the mode, the geometric mean, the harmonic mean, and the median

It must be reported that, even as a book on descriptive statistics, it leaves much to be desired. The errors usually found in such books are to be found here as well as many more. There is the long discussion of skewness and kurtosis based on the false notion that moments are determined by the nature of the distribution in the neighborhood of the mean. Certain properties of the normal distribution are imputed to all distributions. Erroneous criteria for selecting amongst the many means are given. The universality of the normal distribution seems exaggerated; thus, for example, referring to deviations from regressions (page 364): "Since the quantities $(x_0 - \bar{x}_0)$ are 'errors' we may regularly assume them to be normally distributed." Population parameters and their estimates are confused. The book contains a great many statements (like the final one in the section on the Student distribution quoted above) which are so carelessly written that they have to be counted as errors. Several of the derivations and arguments are also carelessly constructed, an extreme example of this appears on page 206: "Is the mean an unbiased statistic? $M = (x_a + x_b + x_c + \cdots + x_n)/N$. Since the various x's are independent, there are just N degrees of freedom and M is unbiased."

Students will likely have difficulty with this book. There is an air of artificiality because of the omission of any discussion of population distributions and the notion of random sampling. Without any background of this kind it is hard to motivate the presentation, and the various topics become isolated. Moments are defined in terms of sample observations, and population moments are defined merely as the limits of these moments as the sample size becomes infinite. To introduce the mean, the author writes essentially: let us consider the function $f(b) = [\Sigma x^b/N]^{1/b}$. There is no pointing to the middle of a distribu-

tion function, or even a sample, or a histogram. The variance is introduced the same way; one considers the function $\Sigma \mid x_i - x_j \mid^m/(N^2 - N)$. Technical terms are used without definition, for example, in the passage about the mean quoted above, the student suddenly encounters the word "unbiased" without definition or previous discussion and must infer its meaning from the context.

Perhaps the best part of the book are three chapters on correlation and regression. The idea of correlation is here introduced with the discussion of a numerical example, and several other topics are discussed in terms of examples. This part of the book is very exhaustive; every sort of correlation coefficient is discussed as is every sort of correction to such coefficients. But still the writing is careless, and there is some confusion of ideas The worst confusion occurs because the distinction between normal and intraclass correlation is never brought out; the discussion hops back and forth between the two ideas with no hint that they are not the same thing. This part of the book, too, is in the style of statistics of thirty years ago; the emphasis is on correlation coefficients rather than regression coefficients.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Leo A. Aroian of Hunter College has been promoted to an assistant professorship.

Mr. Carl A. Bennett is now with the General Electric Co, Hanford Engineering Project, Richland, Washington, as an engineer in the Statistical Division.

Dr. Arthur B. Brown has been promoted from an Assistant Professor to an Associate Professor of Mathematics at Queens College, Flushing, New York.

Professor Maurice H. Belz has returned to the University of Melbourne, Carlton, Australia after having spent six months in the United States.

Dr. Edward E. Cureton, member of Richardson, Bellows, Henry & Co, Inc., industrial psychologists, is now at the United States Naval Air Station, Pensacola, Florida working on a project with the Navy. The object of this project is to improve ground school training, especially instructor training, in the Naval Air Training Command.

Mr. Eric F. Gardner has accepted an assistant professorship at the School of Education, Syracuse University, Syracuse, New York.

Mr. Leo S. Gunlogson, formerly with the Lumbermens Mutual Casualty Co. at Chicago, is now with the Marketing Services Division, Carrier Corporation, Syracuse, New York.

Dr. Theodore E. Harris has accepted a position with the Douglas Aircraft Co, Santa Monica, California.

Dr. Manuel O. Hizon, a former graduate student in the Mathematics Department, University of Michigan, is now with the Bureau of Banking, Manila, Philippines as Actuary-Examiner

Mr. Julius Lieblein, formerly in the Treasury Department, Washington, D C., has transferred to the Statistical Engineering Laboratory, National Bureau of Standards, where he is working on problems in acceptance sampling and process control.

Mr. Jack Moshman, formerly a tutor of mathematics at Queens College, Flushing, New York, has been appointed to the staff of the Department of Mathematics, University of Tennessee.

Dr. Horace W. Norton, formerly with the U S. Weather Bureau, Washington, D. C as meteorologist, is now at Oak Ridge, Tennessee. His position there is to study the application of statistics to reliability of weighings and analyses in connection with accountability for source and fissionable materials.

Mr. Emil D. Schell of the Bureau of Labor Statistics has been appointed Chief of the Mathematics and Electronic Computor Branch in the Office of the Comptroller, United States Air Forces.

Miss Bernice Scherl, formerly with the Schenley Research Institute, Inc., New York, has accepted a position as Statistician, Shell Oil Co., New York

Dr. Irving E Segal, who has been an assistant at the Institute for Advanced Study at Princeton, New Jersey, has accepted an assistant professorship in the Mathematics Department, University of Chicago.

Miss Rosedith Sitgreaves, assistant statistician in the United States Public Health Service, has returned to her position in Washington after doing advanced study at Columbia University.

Dr. John E. Walsh, who received his doctor's degree in mathematics from Princeton University last October, is now employed by Douglas Aircraft Co., Inc. of Santa Monica, California.

Mr. Winfred P. Wilson, a former graduate student at the University of Michigan, has accepted an assistant professorship at the University of Houston, Houston, Texas in the Department of Mathematics.

Announcement has been received of a new journal, *The British Journal of Psychology, Statistical Section*, which is published by the Council of the British Psychological Society. The editors are Professor Sir Cyril Burt and Professor Godfrey Thomson. The first issue has been published and later issues will be published as material warrants. Subscriptions and inquiries should be sent to the University of London Press, Ltd., Warwick Square, London, E. C. 4.

Announcement of Navy Department Joint Board of U. S. Civil Service Examiners

Implementing its scientific research and development program both geographically and in new fields of endeavor, the Navy Department is currently expanding three comparatively new, permanent laboratories in California. Heretofore, the Navy Department's scientific centers have been concentrated in the eastern and eastern seaboard areas.

Two of the laboratories have been established as the logical outgrowth of programs carried on by universities during the war. The Naval Ordnance Test Station, China Lake (formerly Inyokern), California, 160 miles from Los Angeles, was originally an activity of the California Institute of Technology. Its present program involves research, development and test work with ordnance equipment and explosives. The Navy Electronics Laboratory at San Diego, California is the outgrowth of work done by the University of California. It is concerned with research, testing and development of electronic control devices, detection equipment, instrumentation equipment and training aids. The Naval Air Missile Test Center at Point Mugu on the coast of California, 60 miles north of Los Angeles, was established when the need for an installation became apparent as the result of the Navy Department's activities on guided missiles. The Test Center's activities are concerned with flight and laboratory testing and evaluation of guided missiles and their components.

Each of the establishments has current need for qualified personnel in a variety of scientific fields to staff its laboratories Recently completed at the Naval

Ordnance Test Station is Michelson Laboratory at a cost of $6,000,000. Many more millions of dollars have been spent in equipment and facilities. Additional construction and facilities are planned for both the Air Missile Test Center and the Electronics Laboratory

The work programs of the laboratories are planned, directed and accomplished under the direction of an outstanding staff of civilian scientists. Extensive use is made of the council method of operation. Constant liasion is maintained with other research organizations, universities, scientific associations, and outstanding authorities throughout the nation.

Professional positions are in the career service of the Federal government under Civil Service laws. Examinations are now open in the three scientific establishments in the following professional fields: Chemist, Mathematician, Metallurgist, Meteorologist, Physicist, Statistician, Scientific Research Administrator and Scientific Staff Assistant.

Examinations are also open in the following branches of the Engineering profession: Aeronautical, Chemical, Civil, Electrical, Electronics, General, Industrial, Material, Mechanical, Metallurgical, Ordnance, Safety and Structural.

Salaries for most of the positions range from $3397 to $9975 per annum Salaries are predicated on the level of ability, knowledge and experience required to effectively discharge the duties of a specific position

Further information may be obtained from the Navy Department Joint Board of U. S. Civil Service Examiners, 1030 East Green Street, Pasadena 1, California.

Reorganization of Philosophy of Science Association

The Philosophy of Science Association has been reorganized with Philipp Frank of Harvard University as President; C West Churchman of Wayne University, Detroit, as Secretary-Treasurer.

The following are members of the Governing Committee: Gustav Bergmann, State University of Iowa, Thomas A. Cowan, Wayne University; Clyde Kluckhohn, Harvard University; Sebastian Littauer, Columbia University; F. S C. Northrop, Yale Yniversity.

The official journal of the Association is the *Philosophy of Science* of which Professor C. West Churchman is Acting Editor. Manuscripts should be sent to the Acting Editor.

Applications for membership may be sent to the Secretary-Treasurer. Dues are $5.00 a year.

The Association encourages the establishment of local groups in the philosophy of science.

Columbia University Conference on Industrial Experimentation.

The School of Engineering of Columbia University in the City of New York announces an Intersession Five-day Intensive Training Conference on Industrial

Experimentation to be offered September 14–18, 1948 by the Department of Industrial Engineering in cooperation with the Department of Mathematical Statistics of the Graduate Faculty of Political Science.

The lecturing will be shared by Professors S. B. Littauer and J. Wolfwitz and a staff of special lecturers drawn from industry.

A descriptive brochure will be ready for mailing in the latter part of July. For further details, interested persons may communicate directly with Professor S. B. Littauer, Department of Industrial Engineering, Columbia University, New York 27, New York

New Members

The following persons have been elected to membership in the Institute

(December 1, 1947 to February 28, 1948)

Angulo, Walter J., B E. (Johns Hopkins Univ.) Graduate student at Johns Hopkins University, *5229 Beaufort Ave , Baltimore 15, Maryland.*

Beard, Helen P., Ph D. (Mass Institute of Tech.) Assistant Professor of Mathematics, Newcomb College, New Orleans 18, Louisiana.

Blomquist, Nils G., (Univ. of Stockholm) Statistician, Sverige Reinsurance Company, Aladdinsvagen 47, Smedslatten, Sweden.

Bodwell, Charles A., M.S (Univ of Michigan) Graduate student at the University of Michigan, Box 773, West Lodge, Ypsilanti, Michigan.

Burnett, Jean, M.S. (Mich. State College) Instructor in Mathematics, Michigan State College, *702 Cherry Lane, East Lansing, Michigan.*

Burton, Robert E., Student at Michigan University, *1239 Atkinson Avenue, Detroit 2, Michigan*

Byrd, Paul F., M S. (Univ of Chicago) Meteorologist, U.S.A.F , Weather Detachment, Lockbourne Air Base, Columbus 17, Ohio

Cernuschi, Felix, Ph.D (Univ. of Cambridge) Professor at the University of Montevideo, Asociacion Uruguaya de Estadistica, Av. Agraciada 1464, Montevideo, Uruguay.

Connor, William S., Jr., M A (Univ. of North Carolina) Associate Professor of Economics, University of Kentucky, College of Commerce, Lexington, Kentucky

Dalenius, Tore, Fil kand. Hastholmsvagen 16, Stockholm, Sweden.

Davis, Roderic C., M S (Calif Institute of Tech.) P-6 Mathematician, Head of Assessment Section, P O. Box N-467, N O.T.S., Inyokern, California.

Dvoretzky, Aryeh, Ph.D. (Hebrew Univ , Jerusalem) Research Fellow at Hebrew University, Jerusalem, *℅ American Friends Hebrew University, 9 East 89th St., New York.*

Gardner, Robert S., M S (Tulane Univ) Instructor in Statistics, Mathematics Department at Ohio State University, *215 W. Eleventh Avenue, Columbus, Ohio.*

Goins, Mary, M S. (Univ of Mich.) Assistant Professor of Mathematics, Marshall College, *436 Ninth Avenue, Huntington, West Virginia.*

Hratz, Joseph A., B.A. (St. Ambrose College) Instructor of Mathematics, St. Ambrose College, Davenport, Iowa.

Kelly, Harriet J., Ph D (Univ. of Iowa) Research Associate, Head of Statistical Dept. Children's Fund of Michigan, *660 Frederick, Detroit 2, Michigan*

Kincaid, Wilfred M., Ph D (Brown Univ) Instructor in Mathematics, University of Michigan, Ann Arbor, Michigan

Kish, Leslie, B S. (College of the City of N Y.) Senior Sampling Statistician, *701 Mt Pleasant, Ann Arbor, Michigan*

Lehr, Marguerite, Ph D (Bryn Mawr, Pa) Associate Professor of Mathematics, Bryn Mawr College, Cartref, Bryn Mawr, Pennsylvania.

Loizelier, Enrique Blanco, M.A. (Madrid Univ.) Professor of Statistics, Faculty of Economics, Madrid, University, ***Nervion 4, Madrid, Spain.***

Lorenzo, Cesar M., M A. (American Univ , Wash., D C.) Statistician, Food and Agriculture Organization of the United Nations, ***1735 De Sales Street, N.W , Washington 6, D. C.***

Lott, Fred W., Jr., M.A. (Univ. of Mich.) Teaching Fellow at the University of Michigan, ***1233 Malden Court, Willow Run, Michigan.***

Mantel, Nathan, B.S. (City College of New York) Biostatistician, U. S Public Health Service, ***4421 Lily Ponds Dr , N.E., Washington 19, D. C.***

Marrian, Dixon M., A.M (Columbia Univ.) Master at the Gilman Country School, Baltimore, Md., ***1506 Shadyside Road, Baltimore 18, Maryland.***

Martins, Octavio Augusto L., M.A. (Columbia Univ.) Tecnico de Educacao, Department of National Education, Rio de Janeiro, ***Rua Figueiredo de Magalhaes 27, Apt. 702, Rio de Janeiro, Brazil.***

Meyer, Herbert Albert, Ph.D. (Univ. of Iowa) Associate Professor of Mathematics, ***2015 West Columbia, Gainesville, Florida.***

Nikitich, Nicholas, B.C.S. (New York Univ.) Timekeeper, ***20 Featherbed Lane, New York 52, New York.***

Oakland, Gail Barker, M.A. (Univ of Minnesota) Associate Professor of Statistics, University of Manitoba, Winnipeg, Canada

Olkin, Ingram, B.S. (College of City of N Y.) Graduate Student at Columbia University, ***245 Fort Washington Ave , New York 32, New York.***

REPORT ON THE NEW YORK MEETING OF THE INSTITUTE

The thirty-third meeting of the Institute of Mathematical Statistics was held at Columbia University, New York City, New York on Wednesday afternoon and Thursday, April 14 and 15, 1948. The meeting was attended by 158 persons, including the following 78 members of the Institute:

M. Afzal, L. A. Aroian, R. M. Auer, W D Baten, R. E. Bechhofer, J. H. Bushey, J. M. Cameron, B. H Camp, G. C Campbell, S D. Canter, Manuel Cynamon, Tore Dalenius, J. F. Daly, J L. Doob, C. W. Dunnett, Aryeh Dvoretzky, Churchill Eisenhart, Benjamin Epstein, M. W. Eudey, D A. Fraser, M A Geisler, Mary Goins, H. H. Goode, E J. Gumbel, M. H Hansen, Mina Haskind, L H. Herback, S M Ikhtiar-ul-Mulk, Seymour Jablon, L. F. Knudsen, Jack Laderman, Howard Levene, S. B Littauer, F. M. Lord, Irving Lorge, Eugene Lukacs, W. G Madow, Sophie Marcuse, Robert Mirsky, E. B. Mode, D J Morrow, Frederick Mosteller, D. N Nanda, P M. Neurath, G. E Noether, M L Norden, Ingram Olkin, P S. Olmstead, A. L. O'Toole, Katharine Pease, E J. Pitman, W. A Reynolds, J S Rhodes, H. E. Robbins, H G. Romig, Ernest Rubin, Herman Rubin, P. J. Rulon, Frank Saidel, G. R Seth, M. A. Schlorok, S. S. Shrikhande, Rosedith Sitgreaves, Milton Sobel, Emma Spaney, F. F. Stephan, B. R. Suydam, Henry Teicher, J. W Tukey, A. Wald, H M. Walker, J. E. Walsh, S. S. Wilks, Dzung-shu Wei, Lionel Weiss, Jacob Wolfowitz, C A Wright, Mohammad Yusuf

The Wednesday afternoon session, Professor S. B. Littauer of Columbia University presiding, was devoted to the following two invited addresses:

1 *Incomplete Block Designs*
   Professor R C. Bose, Calcutta University and the University of North Carolina
2. *Non-Parametric Inference*
   Professor J. G Pitman, University of Tasmania and Columbia University

The Thursday morning session, Professor Hobart Bushey of Hunter College presiding, consisted of a Symposium on *Scales of Measurement* at which two invited papers·

1 *The Development of Psychological Scaling Techniques*
   Professor Harold Gulliksen, Princeton University
2 *A Generalized Model for Scales*
   Professor Paul Lazarsfeld, Columbia University

were followed by prepared discussion by Professors Phillip Rulon of Harvard University and John Tukey of Princeton University.

The Thursday afternoon session, Dr. Harry G. Romig of Bell Telephone Laboratories presiding, was devoted to the following contributed papers:

1 *Optimum Character of the Sequential Probability Ratio Test*
   Professors Abraham Wald and Jacob Wolfowitz, Columbia University
2. *Multi-parameter Sequential Estimation*
   Mr G R. Seth, Columbia University

3. *The Distribution of a Definite Quadratic Form*
   Professor Herbert Robbins, University of North Carolina
4. *The Moments and Cumulants of the Product of 2, 3, or 4 Dependent Variables (Preliminary Report)*
   Professor Leo A Aroian, Hunter College
5. *Generalization to N Dimensions of Inequalities of the Tchebycheff Type*
   Professor Burton H. Camp, Wesleyan University
6. *On the Power Function of a Sign Test Formed by Using Subsamples*
   Dr. John E. Walsh, Project Rand
7. *The Distribution of T^2, a Multivariate Generalization of the F-test*
   Miss Dorothy J Morrow, University of North Carolina.
8. *Approximate Confidence Points (Preliminary Report)*
   Professor John Tukey, Princeton University

At all of the sessions there was active discussion from the floor.

On Wednesday evening members and guests had dinner at the Men's Faculty Club.

S. B. Littauer
Assistant Secretary

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE
OF MATHEMATICAL STATISTICS

Contents

Vol. XIX, No. 3 — September, 1948

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves, Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, P. S. Dwyer, 116 Rackham Hall, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $8.00 inside the Western Hemisphere and $5.00 elsewhere. Single copies $3.00. Back numbers are available at $8.00 per volume or $3.00 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879

A CLASS OF STATISTICS WITH ASYMPTOTICALLY NORMAL DISTRIBUTION[1]

By Wassily Hoeffding

Institute of Statistics, University of North Carolina

1. Summary. Let $X_1, \cdots, X_n$ be n independent random vectors, $X_\nu = (X_\nu^{(1)}, \cdots, X_\nu^{(r)})$, and $\Phi(x_1, \cdots, x_m)$ a function of $m(\leq n)$ vectors $x_\nu = (x_\nu^{(1)}, \cdots, x_\nu^{(r)})$. A statistic of the form $U = \Sigma''\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m})/n(n-1) \cdots (n-m+1)$, where the sum Σ'' is extended over all permutations $(\alpha_1, \cdots, \alpha_m)$ of m different integers, $1 \leq \alpha_i \leq n$, is called a U-statistic. If $X_1, \cdots, X_n$ have the same (cumulative) distribution function (d.f.) $F(x)$, U is an unbiased estimate of the population characteristic $\theta(F) = \int \cdots \int \Phi(x_1, \cdots, x_m) dF(x_1) \cdots dF(x_m)$. $\theta(F)$ is called a regular functional of the d.f. $F(x)$. Certain optimal properties of U-statistics as unbiased estimates of regular functionals have been established by Halmos [9] (cf. Section 4).

The variance of a U-statistic as a function of the sample size n and of certain population characteristics is studied in Section 5.

It is shown that if $X_1, \cdots, X_n$ have the same distribution and $\Phi(x_1, \cdots, x_m)$ is independent of n, the d.f. of $\sqrt{n}(U - \theta)$ tends to a normal d.f. as $n \to \infty$ under the sole condition of the existence of $E\Phi^2(X_1, \cdots, X_m)$. Similar results hold for the joint distribution of several U-statistics (Theorems 7.1 and 7.2), for statistics U' which, in a certain sense, are asymptotically equivalent to U (Theorems 7.3 and 7.4), for certain functions of statistics U or U' (Theorem 7.5) and, under certain additional assumptions, for the case of the X_ν's having different distributions (Theorems 8.1 and 8.2). Results of a similar character, though under different assumptions, are contained in a recent paper by von Mises [18] (cf. Section 7).

Examples of statistics of the form U or U' are the moments, Fisher's k-statistics, Gini's mean difference, and several rank correlation statistics such as Spearman's rank correlation and the difference sign correlation (cf. Section 9). Asymptotic power functions for the non-parametric tests of independence based on these rank statistics are obtained. They show that these tests are not unbiased in the limit (Section 9f). The asymptotic distribution of the coefficient of partial difference sign correlation which has been suggested by Kendall also is obtained (Section 9h).

2. Functionals of distribution functions. Let $F(x) = F(x^{(1)}, \cdots, x^{(r)})$ be an r-variate d.f. If to any F belonging to a subset $\mathfrak{D}$ of the set of all d.f.'s in the r-dimensional Euclidean space is assigned a quantity $\theta(F)$, then $\theta(F)$ is called a

[1] Research under a contract with the Office of Naval Research for development of multivariate statistical theory.

functional of F, defined on $\mathfrak{D}$. In this paper the word functional will always mean functional of a d f.

An infinite population may be considered as completely determined by its d f, and any numerical characteristic of an infinite population with d.f. F that is used in statistics is a functional of F. A finite population, or sample, of size n is determined by its d.f., $S(x)$ say, and its size n. n itself is not a functional of S since two samples of different size may have the same d.f

If $S(x^{(1)}, \cdots, x^{(r)})$ is the d.f. of a finite population, or a sample, consisting of n elements

$$x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n), \tag{2.1}$$

then $nS(x^{(1)}, \cdots, x^{(r)})$ is the number of elements x_α such that

$$x_\alpha^{(1)} \leq x^{(1)}, \cdots, x_\alpha^{(r)} \leq x^{(r)}$$

Since $S(x^{(1)}, \cdots, x^{(r)})$ is symmetric in $x_1, \cdots, x_n$, and retains its value for a sample formed from the sample (2.1) by adding one or more identical samples, the same two properties hold true for a sample functional $\theta(S)$. Most statistics in current use are functions of n and of functionals of the sample d f

A random sample $\{X_1, \cdots, X_n\}$ is a set of n independent random vectors

$$X_\alpha = (X_\alpha^{(1)}, \cdots, X_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n). \tag{2.2}$$

For any fixed values $x^{(1)}, \cdots, x^{(r)}$, the d.f. $S(x^{(1)}, \cdots, x^{(r)})$ of a random sample is a random variable. The functional $\theta(S)$, where S is the d f of the random sample, is itself a random variable, and may be called a random functional.

A remarkable application of the theory of functionals to functionals of d.f 's has been made by von Mises [18] who considers the asymptotic distributions of certain functionals of sample d f 's (Cf also Section 7)

3. Unbiased estimation and regular functionals. Consider a functional $\theta = \theta(F)$ of the r-variate d.f $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and suppose that for some sample size n, θ admits an unbiased estimate for any d.f F in $\mathfrak{D}$ That is, if $X_1, \cdots, X_n$ are n independent random vectors with the same d f. F, there exists a function $\varphi(x_1, \cdots, x_n)$ of n vector arguments (2.1) such that the expected value of $\varphi(X_1, \cdots, X_n)$ is equal to $\theta(F)$, or

$$\int \cdots \int \varphi(x_1, \cdots, x_n)\, dF(x_1) \cdots dF(x_n) = \theta(F) \tag{3.1}$$

for every F in $\mathfrak{D}$ Here and in the sequel, when no integration limits are indicated, the integral is extended over the entire space of $x_1, \cdots, x_n$. The integral is understood in the sense of Stieltjes-Lebesgue.

The estimate $\varphi(x_1, \cdots, x_n)$ of $\theta(F)$ is called unbiased over $\mathfrak{D}$.

A functional $\theta(F)$ of the form (3.1) will be referred to as *regular over* $\mathfrak{D}$.[2]

[1] This is an adaptation to functionals of d f.'s of the term "regular functional" used by Volterra [21].

Thus, the functionals regular over $\mathfrak{D}$ are those admitting an unbiased estimate over $\mathfrak{D}$

If $\theta(F)$ is regular over $\mathfrak{D}$, let $m(\leq n)$ be the smallest sample size for which there exists an unbiased estimate $\Phi(x_1, \cdots, x_m)$ of θ over $\mathfrak{D}$:

$$\theta(F) = \int \cdots \int \Phi(x_1, \cdots, x_m)\, dF(x_1) \cdots dF(x_m) \tag{3 2}$$

for any F in $\mathfrak{D}$. Then m will be called *the degree over* $\mathfrak{D}$ of the regular functional $\theta(F)$.

If the expected value of $\varphi(X_1, \cdots, X_n)$ is equal to $\theta(F)$ whenever it exists, $\varphi(x_1, \cdots, x_n)$ will be called a *distribution-free unbiased estimate* (d-f.u.e.) of $\theta(F)$. The degree of $\theta(F)$ over the set $\mathfrak{D}_0$ of d.f.'s F for which the right hand side of (3.1) exists will be simply termed the *degree* of $\theta(F)$.

A regular functional of degree 1 over $\mathfrak{D}$ is called a linear regular functional over $\mathfrak{D}$ If $\theta(F)$ has the same value for all F in $\mathfrak{D}$, $\theta(F)$ may be termed a regular functional of degree zero over $\mathfrak{D}$

Any function $\Phi(x_1, \cdots, x_m)$ satisfying (3.2) will be referred to as a *kernel* of the regular functional $\theta(F)$

For any regular functional $\theta(F)$ there exists a kernel $\Phi_0(x_1, \cdots, x_m)$ symmetric in $x_1, \cdots, x_m$. For if $\Phi(x_1, \cdots, x_m)$ is a kernel of $\theta(F)$,

$$\Phi_0(x_1, \cdots, x_m) = \frac{1}{m!} \Sigma \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}), \tag{3 3}$$

where the sum is taken over all permutations $(\alpha_1, \cdots, \alpha_m)$ of $(1, \cdots, m)$, is a symmetric kernel of $\theta(F)$.

If $\theta_1(F)$ and $\theta_2(F)$ are two regular functionals of degrees m_1 and m_2 over $\mathfrak{D}$, then the sum $\theta_1(F) + \theta_2(F)$ and the product $\theta_1(F)\theta_2(F)$ are regular functionals of degrees $\leq m = \text{Max}(m_1, m_2)$ and $\leq m_1 + m_2$, respectively, over $\mathfrak{D}$. For if $\Phi_i(x_1, \cdots, x_{m_i})$ is a kernel of $\theta_i(F)$, $(i = 1, 2)$, then

$$\theta_1(F) + \theta_2(F) = \int \cdots \int \{\Phi_1(x_1, \cdots, x_{m_1}) + \Phi_2(x_1, \cdots, x_{m_2})\} \, dF(x_1) \cdots dF(x_m)$$

and

$$\theta_1(F)\theta_2(F) = \int \cdots \int \Phi_1(x_1, \cdots, x_{m_1})\Phi_2(x_{m_1+1}, \cdots, x_{m_1+m_2}) \, dF(x_1) \cdots dF(x_{m_1+m_2})$$

More generally, *a polynomial in regular functionals is itself a regular functional*

Examples of linear regular functionals are the moments about the origin,

$$\mu'_{\nu_1, \cdots, \nu_r} = \int \cdots \int (x^{(1)})^{\nu_1} \cdots (x^{(r)})^{\nu_r} \, dF(x^{(1)}, \cdots, x^{(r)}).$$

A moment about the mean is a polynomial in moments μ' about 0, and hence a regular functional over the set $\mathfrak{D}_0$ of d.f.'s for which it exists (cf. Halmos [9]). For instance, the variance of $X^{(1)}$,

$$\sigma^2 = \int\int ((x_1^{(1)})^2 - x_1^{(1)} x_2^{(1)}) \, dF(x_1^{(1)}) \, dF(x_2^{(1)})$$

is a regular functional of degree 2. A symmetrical kernel of σ^2 is $(x^{(1)} - x^{(2)})^2/2$. If $\mathfrak{D}$ is the set of univariate d.f.'s with mean μ and existing second moment, σ^2 is a linear regular functional of F over $\mathfrak{D}$, since then we have

$$\sigma^2 = \int (x_1^{(1)} - \mu)^2 \, dF(x_1^{(1)}).$$

The function

$$v = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} \frac{1}{2} (x_\alpha^{(1)} - x_\beta^{(1)})^2 = \frac{1}{n-1} \sum_\alpha \left(x_\alpha^{(1)} - \frac{1}{n} \sum_\beta x_\beta^{(1)} \right)^2$$

is a distribution-free unbiased estimate of σ^2. The function

$$\Gamma\left(\frac{n-1}{2}\right) \sqrt{\frac{n-1}{2}} \sqrt{v} / \Gamma\left(\frac{n}{2}\right)$$

is known to be an unbiased estimate of σ over the set of univariate normal d.f.'s, but it is not a d.-f. u.e.

4. *U*-statistics. Let $x_1, \cdots, x_n$ be a sample of n vectors (2.1) and $\Phi(x_1, \cdots, x_m)$ a function of $m(\leq n)$ vector arguments. Consider the function of the sample,

$$(4.1) \quad U = U(x_1, \cdots, x_n) = \frac{1}{n(n-1) \cdots (n-m+1)} \Sigma'' \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}),$$

where Σ'' stands for summation over all permutations $(\alpha_1, \cdots, \alpha_m)$ of m integers such that

$$(4.2) \qquad 1 \leq \alpha_i \leq n, \qquad \alpha_i \neq \alpha_j \text{ if } i \neq j, \qquad (i, j = 1, \cdots, m)$$

U is the average of the values of Φ in the set of ordered subsets of m members of the sample (2.1). U is symmetric in $x_1, \cdots, x_n$.

Any statistic of the form (4.1) will be called a *U-statistic*. Any function $\Phi(x_1, \cdots, x_m)$ satisfying (4.1) will be referred to as a *kernel* of the statistic U.

If $\Phi(x_1, \cdots, x_m)$ is a kernel of a regular functional $\theta(F)$ defined on a set $\mathfrak{D}$, then U is an unbiased estimate of $\theta(F)$ over $\mathfrak{D}$:

$$(4.3) \qquad \theta(F) = \int \cdots \int U(x_1, \cdots, x_n) \, dF(x_1) \cdots dF(x_n)$$

for every F in $\mathfrak{D}$.

For $n = m$, U reduces to the symmetric kernel (3.3) of $\theta(F)$

From a recent paper by Halmos [9] it follows for the case of univariate d f.'s ($r = 1$):

If $\theta(F)$ is a regular functional of degree m over a set $\mathfrak{D}$ containing all purely discontinuous d f 's, U is the only unbiased estimate over $\mathfrak{D}$ which is symmetric in $x_1, \cdots, x_n$, and U has the least variance among all unbiased estimates over $\mathfrak{D}$.

These results and the proofs given by Halmos can easily be extended to the multivariate case ($r > 1$).

Combining (3 3) and (4.1) we may write a U-statistic in the form

$$U(x_1, \cdots, x_n) = \binom{n}{m}^{-1} \Sigma' \Phi_0(x_{\alpha_1}, \cdots, x_{\alpha_m}), \tag{4.4}$$

where the kernel Φ_0 is symmetric in its m vector arguments and the sum Σ' is extended over all subscripts α such that

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_m \leq n.$$

Another statistic frequently used for estimating $\theta(F)$ is $\theta(S)$, where $S = S(x)$ is the d.f. of the sample (2 1). If S is substituted for F in (3.2), we have

$$\theta(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_m=1}^{n} \Phi(x_{\alpha_1}, \cdots, x_{\alpha_m}). \tag{4.5}$$

In particular, the sample moments have this form; their kernel Φ is obtained by the method described in section 3

If $m = 1$, $\theta(S) = U$. If $m = 2$,

$$\theta(S) = \frac{n-1}{n} U + \frac{1}{n}\left\{\frac{1}{n} \sum_{\alpha=1}^{n} \Phi(x_\alpha, x_\alpha)\right\},$$

and $\theta(S)$ is a linear function of U-statistics with coefficients depending on n. This is easily seen to be true for any m. In general $\theta(S)$ is not an unbiased estimate of $\theta(F)$ If, however, the expected value of $\theta(S)$ exists for every F in $\mathfrak{D}$, we have

$$E\{\theta(S)\} = \theta(F) + O(n^{-1}),$$

and the estimate $\theta(S)$ of $\theta(F)$ may be termed unbiased in the limit over $\mathfrak{D}$.

Numerous statistics in current use have the form of, or can be expressed in terms of U-statistics. From what was said above about moments as regular functionals, it is easy to obtain U-statistics which are d.-f. u e.'s of the moments about the mean of any order (cf. Halmos [9]) Fisher's k-statistics are U-statistics, as follows from their definition as unbiased estimates of the cumulants, symmetric in the sample values Another example is Gini's mean difference

$$\frac{1}{n(n-1)} \sum_{\alpha \neq \beta} | x_\alpha^{(1)} - x_\beta^{(1)} |.$$

More examples, in particular of rank correlation statistics, will be given in section 9.

5. The variance of a U-statistic. Let $X_1, \cdots, X_n$ be n independent random vectors with the same d.f. $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and let

$$(5.1) \qquad U = U(X_1, \cdots, X_n) = \binom{n}{m}^{-1} \Sigma' \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}),$$

where $\Phi(x_1, \cdots, x_m)$ is symmetric in $x_1, \cdots, x_m$ and Σ' has the same meaning as in (4.4). Suppose that the function Φ does not involve n.

If $\theta = \theta(F)$ is defined by (3.2), we have

$$E\{U\} = E\{\Phi(X_1, \cdots, X_m)\} = \theta.$$

Let

$$(5.2) \qquad \Phi_c(x_1, \cdots, x_c) = E\{\Phi(x_1, \cdots, x_c, X_{c+1}, \cdots, X_m)\}, \quad (c = 1, \cdots, m),$$

where $x_1, \cdots, x_c$ are arbitrary fixed vectors and the expected value is taken with respect to the random vectors $X_{c+1}, \cdots, X_m$. Then

$$(5.3) \qquad \Phi_{c-1}(x_1, \cdots, x_{c-1}) = E\{\Phi_c(x_1, \cdots, x_{c-1}, X_c)\},$$

and

$$(5.4) \qquad E\{\Phi_c(X_1, \cdots, X_c)\} = \theta, \qquad (c = 1, \cdots, m).$$

Define

$$(5.5) \qquad \Psi(x_1, \cdots, x_m) = \Phi(x_1, \cdots, x_m) - \theta,$$

$$(5.6) \qquad \Psi_c(x_1, \cdots, x_c) = \Phi_c(x_1, \cdots, x_c) - \theta, \qquad (c = 1, \cdots, m).$$

We have

$$(5.7) \qquad \Psi_{c-1}(x_1, \cdots, x_{c-1}) = E\{\Psi_c(x_1, \cdots, x_{c-1}, X_c)\},$$

$$(5.8) \qquad E\{\Psi_c(X_1, \cdots, X_c)\} = E\{\Psi(X_1, \cdots, X_m)\} = 0, \qquad (c = 1, \cdots, m).$$

Suppose that the variance of $\Psi_c(X_1, \cdots, X_c)$ exists, and let

$$(5.9) \qquad \zeta_0 = 0, \qquad \zeta_c = E\{\Psi_c^2(X_1, \cdots, X_c)\}, \qquad (c = 1, \cdots, m).$$

We have

$$(5.10) \qquad \zeta_c = E\{\Phi_c^2(X_1, \cdots, X_c)\} - \theta^2.$$

$\zeta_c = \zeta_c(F)$ is a polynomial in regular functionals of F, and hence itself a regular functional of F (of degree $\leq 2m$).

If, for some parent distribution $F = F_0$ and some integer d, we have $\zeta_d(F_0) = 0$, this means that $\Psi_d(X_1, \cdots, X_d) = 0$ with probability 1. By (5.7) and (5.9), $\zeta_d = 0$ implies $\zeta_1 = \cdots = \zeta_{d-1} = 0$.

If $\zeta_1(F_0) = 0$, we shall say that the regular functional $\theta(F)$ is *stationary*[3] for $F = F_0$. If

$$\zeta_1(F_0) = \cdots = \zeta_d(F_0) = 0, \qquad \zeta_{d+1}(F_0) > 0, \qquad (1 \leq d \leq m), \tag{5.11}$$

$\theta(F)$ will be called *stationary of order d* for $F = F_0$.

If $(\alpha_1, \cdots, \alpha_m)$ and $(\beta_1, \cdots, \beta_m)$ are two sets of m different integers, $1 \leq \alpha_i$, $\beta_i \leq n$, and c is the number of integers common to the two sets, we have, by the symmetry of Ψ,

$$E\{\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\Psi(X_{\beta_1}, \cdots, X_{\beta_m})\} = \zeta_c \tag{5.12}$$

If the variance of U exists, it is equal to

$$\sigma^2(U) = \binom{n}{m}^{-2} E\{\Sigma'\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\}^2$$
$$= \binom{n}{m}^{-2} \sum_{c=0}^{m} \Sigma^{(c)} E\{\Psi(X_{\alpha_1}, \cdots, X_{\alpha_m})\Psi(X_{\beta_1}, \cdots, X_{\beta_m})\},$$

where $\Sigma^{(c)}$ stands for summation over all subscripts such that

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_m \leq n, \qquad 1 \leq \beta_1 < \beta_2 < \cdots < \beta_m \leq n,$$

and exactly c equations

$$\alpha_i = \beta_j$$

are satisfied. By (5.12), each term in $\Sigma^{(c)}$ is equal to ζ_c. The number of terms in $\Sigma^{(c)}$ is easily seen to be

$$\frac{n(n-1)\cdots(n-2m+c+1)}{c!(m-c)!(m-c)!} = \binom{m}{c}\binom{n-m}{m-c}\binom{n}{m},$$

and hence, since $\zeta_0 = 0$,

$$\sigma^2(U) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c}\zeta_c \tag{5.13}$$

When the distributions of $X_1, \cdots, X_n$ are different, $F_\nu(x)$ being the d.f. of X_ν, let

$$\theta_{\alpha_1, \cdots, \alpha_m} = E\{\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m})\}, \tag{5.14}$$

$$\Psi_{c(\alpha_1, \cdots, \alpha_c)\beta_1, \cdots, \beta_{m-c}}(x_1, \cdots, x_c)$$
$$= E\{\Phi(x_1, \cdots, x_c, X_{\beta_1}, \cdots, X_{\beta_{m-c}})\} - \theta_{\alpha_1, \cdots, \alpha_c, \beta_1, \cdots, \beta_{m-c}}, \tag{5.15}$$
$$(c = 1, \cdots, m),$$

---

[3] According to the definition of the derivative of a functional (cf. Volterra [21]; for functionals of d.f.'s cf. von Mises [18]), the function $m(m-1)\cdots(m-d+1)\Psi_d(x_1, \cdots, x_d)$, which is a functional of F, is a d-th derivative of $\theta(F)$ with respect to F at the "point" F of the space of d.f.'s.

$$\zeta_{c(\alpha_1, \cdots, \alpha_c)\beta_1, \cdots, \beta_{m-c}, \gamma_1, \cdots, \gamma_{m-c}}$$

$$(5.16) \qquad = E\{\Psi_{c(\alpha_1, \cdots, \alpha_c)\beta_1, \cdots, \beta_{m-c}}(X_{\alpha_1}, \cdots, X_{\alpha_c})\Psi_{c(\alpha_1, \cdots, \alpha_c)\gamma_1, \cdots, \gamma_{m-c}}(X_{\alpha_1}, \cdots, X_{\alpha_c})\}$$

$$(5.17) \qquad \zeta_{c,n} = \frac{c!(m-c)!(m-c)!}{n(n-1)\cdots(n-2m+c+1)} \sum \zeta_{c(\alpha_1, \cdots, \alpha_c)\beta_1, \cdots, \beta_{m-c}, \gamma_1, \cdots, \gamma_{m-c}}$$

where the sum is extended over all subscripts α, β, γ such that

$$1 \le \alpha_1 < \cdots < \alpha_c \le n, \quad 1 \le \beta_1 < \cdots < \beta_{m-c} \le n, \quad 1 \le \gamma_1 < \cdots \gamma_{m-c} \le n,$$

$$\alpha_i \ne \beta_j, \qquad \alpha_i \ne \gamma_j, \qquad \beta_i \ne \gamma_j .$$

Then the variance of U is equal to

$$(5.18) \qquad \sigma^2(U) = \binom{n}{m}^{-1} \sum_{c=1}^{m} \binom{m}{c}\binom{n-m}{m-c} \zeta_{c,n} .$$

Returning to the case of identically distributed X's, we shall now prove some inequalities satisfied by $\zeta_1, \cdots, \zeta_m$ and $\sigma^2(U)$ which are contained in the following theorems:

THEOREM 5.1 *The quantities $\zeta_1, \cdots, \zeta_m$ as defined by (5.9) satisfy the inequalities*

$$(5.19) \qquad 0 \le \frac{\zeta_c}{c} \le \frac{\zeta_d}{d} \qquad \text{if } 1 \le c < d \le m.$$

THEOREM 5.2 *The variance $\sigma^2(U_n)$ of a U-statistic $U_n = U(X_1, \cdots, X_n)$, where $X_1, \cdots, X_n$ are independent and identically distributed, satisfies the inequalities*

$$(5.20) \qquad \frac{m^2}{n} \zeta_1 \le \sigma^2(U_n) \le \frac{m}{n} \zeta_m .$$

$n\sigma^2(U_n)$ is a decreasing function of n,

$$(5.21) \qquad (n+1)\sigma^2(U_{n+1}) \le n\sigma^2(U_n),$$

which takes on its upper bound $m\zeta_m$ for $n = m$ and tends to its lower bound $m^2\zeta_1$ as n increases:

$$(5.22) \qquad \sigma^2(U_m) = \zeta_m ,$$

$$(5.23) \qquad \lim_{n\to\infty} n\sigma^2(U_n) = m^2 \zeta_1 .$$

If $E\{U_n\} = \theta(F)$ is stationary of order $\ge d-1$ for the d.f. of X_α, (5.20) may be replaced by

$$(5.24) \qquad \frac{m}{d} K_n(m, d)\zeta_d \le \sigma^2(U_n) \le K_n(m, d)\zeta_m ,$$

where

$$K_n(m, d) = \binom{n}{m}^{-1} \sum_{c=d}^{m} \binom{m-1}{c-1}\binom{n-m}{m-c}. \tag{5.25}$$

We postpone the proofs of Theorems 5.1 and 5.2.

(5.13) and (5.19) imply that a necessary and sufficient condition for the existence of $\sigma^2(U)$ is the existence of

$$\zeta_m = E\{\Phi^2(X_1, \cdots, X_m)\} - \theta^2 \tag{5.26}$$

or that of $E\{\Phi^2(X_1, \cdots, X_m)\}$

If $\zeta_1 > 0$, $\sigma^2(U)$ is of order n^{-1}

If $\theta(F)$ is stationary of order d for $F = F_0$, that is, if (5 11) is satisfied, $\sigma^2(U)$ is of order n^{-d-1} Only if, for some $F = F_0$, $\theta(F)$ is stationary of order m, where m is the degree of $\theta(F)$, we have $\sigma^2(U) = 0$, and U is equal to a constant with probability 1

For instance, if $\theta(F_0) = 0$, the functional $\theta^2(F)$ is stationary for $F = F_0$. Other examples of stationary "points" of a functional will be found in section 9d

For proving Theorem 5.1 we shall require the following

LEMMA 5 1 *If*

$$\delta_d = \zeta_d - \binom{d}{1}\zeta_{d-1} + \binom{d}{2}\zeta_{d-2} \cdots + (-1)^{d-1}\binom{d}{d-1}\zeta_1, \tag{5 27}$$

we have

$$\delta_d \geq 0, \qquad (d = 1, \cdots, m) \tag{5.28}$$

and

$$\zeta_d = \delta_d + \binom{d}{1}\delta_{d-1} + \cdots + \binom{d}{d-1}\delta_1. \tag{5 29}$$

PROOF (5 29) follows from (5 27) by induction

For proving (5 28) let

$$\eta_0 = \theta^2, \qquad \eta_c = E\{\Phi_c^2(X_1, \cdots, X_c)\}, \qquad (c = 1, \cdots, m).$$

Then, by (5 10),

$$\zeta_c = \eta_c - \eta_0,$$

and on substituting this in (5.27) we have

$$\delta_d = \sum_{c=0}^{d} (-1)^{d-c}\binom{d}{c}\eta_c$$

From (5.9) it is seen that (5.28) is true for $d = 1$. Suppose that (5 28) holds for $1, \cdots, d-1$. Then (5.28) will be shown to hold for d.

Let

$$\bar{\Phi}_0(x_1) = \Phi_1(x_1) - \theta, \qquad \bar{\Phi}_c(x_1, x_2, \cdots, x_{c+1})$$
$$= \Phi_{c+1}(x_1, \cdots, x_{c+1}) - \Phi_c(x_2, \cdots, x_{c+1}), \qquad (c = 1, \cdots, d-1).$$

For an arbitrary fixed x_1, let

$$\bar{\eta}_c(x_1) = E\{\bar{\Phi}_c^2(x_1, X_2, \cdots, X_{c+1})\}, \qquad (c = 0, \cdots, d-1).$$

Then, by induction hypothesis,

$$\bar{\delta}_{d-1}(x_1) = \sum_{c=0}^{d-1} (-1)^{d-1-c} \binom{d-1}{c} \bar{\eta}_c(x_1) \geq 0$$

for any fixed x_1.

Now,

$$E\{\bar{\eta}_c(X_1)\} = \eta_{c+1} - \eta_c,$$

and hence

$$E\{\bar{\delta}_{d-1}(X_1)\} = \sum_{c=0}^{d-1} (-1)^{d-1-c}\binom{d-1}{c}(\eta_{c+1} - \eta_c) = \sum_{c=0}^{d} (-1)^{d-c}\binom{d}{c}\eta_c = \delta_d.$$

The proof of Lemma 5 1 is complete.

Proof of Theorem 5 1. By (5 29) we have for $c < d$

$$c\zeta_d - d\zeta_c = c\sum_{a=1}^{d}\binom{d}{a}\delta_a - d\sum_{a=1}^{c}\binom{c}{a}\delta_a$$

$$= \sum_{a=1}^{c}\left[c\binom{d}{a} - d\binom{c}{a}\right]\delta_a + c\sum_{a=c+1}^{d}\binom{d}{a}\delta_a. \tag{5.30}$$

From (5 28), and since $c\binom{d}{a} - d\binom{c}{a} \geq 0$ if $1 \leq a \leq c \leq d$, it follows that each term in the two sums of (5 30) is not negative. This, in connection with (5.9) proves Theorem 5 1.

Proof of Theorem 5 2. From (5 19) we have

$$c\zeta_1 \leq \zeta_c \leq \frac{c}{m}\zeta_m, \qquad (c = 1, \cdots, m).$$

Applying these inequalities to each term in (5.13) and using the identity

$$\binom{n}{m}^{-1}\sum_{c=1}^{m} c\binom{m}{c}\binom{n-m}{m-c} = \frac{m^2}{n}, \tag{5.31}$$

we obtain (5.20)

(5 22) and (5.23) follow immediately from (5 13).

For (5.21) we may write

$$D_n \geq 0, \tag{5.32}$$

where

$$D_n = n\sigma^2(U_n) - (n+1)\sigma^2(U_{n+1})$$

Let

$$D_n = \sum_{c=1}^{m} d_{n,c}\zeta_c.$$

Then we have from (5.13)

$$d_{n,c} = n\binom{m}{c}\binom{n-m}{m-c}\binom{n}{m}^{-1} - (n+1)\binom{m}{c}\binom{n+1-m}{m-c}\binom{n+1}{m}^{-1}, \tag{5.33}$$

or

$$d_{n,c} = \binom{m}{c}\binom{n-m+1}{m-c}(n-m+1)^{-1}\binom{n}{m}^{-1}\{(c-1)n-(m-1)^2\},$$
$$(1 \leq c \leq m \leq n).$$

Putting

$$c_0 = 1 + \left[\frac{(m-1)^2}{n}\right],$$

where $[u]$ denotes the largest integer $\leq u$, we have

$$d_{n,c} \leq 0 \qquad \text{if } c \leq c_0,$$
$$d_{n,c} > 0 \qquad \text{if } c > c_0$$

Hence, by (5 19),

$$d_{n,c}\zeta_c \geq \frac{1}{c_0}\zeta_{c_0} c d_{n,c}, \qquad (c = 1, \cdots, m),$$

and

$$D_n \geq \frac{1}{c_0}\zeta_{c_0}\sum_{c=1}^{m} c d_{n,c}.$$

By (5 33) and (5.31), the latter sum vanishes. This proves (5 32).

For the stationary case $\zeta_1 = \cdot \quad = \zeta_{d-1} = 0$, (5.24) is a direct consequence of (5.13) and (5.19) The proof of Theorem 5.2 is complete

6. The covariance of two U-statistics. Consider a set of g U-statistics,

$$U^{(\gamma)} = \binom{n}{m(\gamma)}^{-1}\Sigma'\Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}), \qquad (\gamma = 1, \cdots, g),$$

each $U^{(\gamma)}$ being a function of the same n independent, identically distributed random vectors $X_1, \cdots, X_n$. The function $\Phi^{(\gamma)}$ is assumed to be symmetric in its $m(\gamma)$ arguments ($\gamma = 1, \cdots, g$).

Let

$$E\{U^{(\gamma)}\} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\} = \theta^{(\gamma)}, \qquad (\gamma = 1, \cdots, g);$$

$$\Psi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}) = \Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}) - \theta^{(\gamma)}, \qquad (\gamma = 1, \cdots, g); \tag{6.1}$$

$$\Psi_c^{(\gamma)}(x_1, \cdots, x_c) = E\{\Psi^{(\gamma)}(x_1, \cdots, x_c, X_{c+1}, \cdots, X_{m(\gamma)})\}, \qquad (c = 1, \cdots, m(\gamma); \gamma = 1, \cdots, g), \tag{6.2}$$

$$\zeta_c^{(\gamma,\delta)} = E\{\Psi_c^{(\gamma)}(X_1, \cdots, X_c)\Psi_c^{(\delta)}(X_1, \cdots, X_c)\}, \qquad (\gamma, \delta = 1, \cdots, g). \tag{6.3}$$

If, in particular, $\gamma = \delta$, we shall write

$$\zeta_c^{(\gamma)} = \zeta_c^{(\gamma,\gamma)} = E\{\Psi_c^{(\gamma)}(X_1, \cdots, X_c)\}^2. \tag{6.4}$$

Let

$$\sigma(U^{(\gamma)}, U^{(\delta)}) = E\{(U^{(\gamma)} - \theta^{(\gamma)})(U^{(\delta)} - \theta^{(\delta)})\}$$

be the covariance of $U^{(\gamma)}$ and $U^{(\delta)}$.

In a similar way as for the variance, we find, if $m(\gamma) \le m(\delta)$,

$$\sigma(U^{(\gamma)}, U^{(\delta)}) = \binom{n}{m(\gamma)}^{-1} \sum_{c=1}^{m(\gamma)} \binom{m(\delta)}{c}\binom{n - m(\delta)}{m(\gamma) - c}\zeta_c^{(\gamma,\delta)}. \tag{6.5}$$

The right hand side is easily seen to be symmetric in γ, δ.

For $\gamma = \delta$, (6.5) is the variance of $U^{(\gamma)}$ (cf (5.13)).

We have from (5.23) and (6.5)

$$\lim_{n\to\infty} n\sigma^2(U^{(\gamma)}) = m^2(\gamma)\zeta_1^{(\gamma)},$$

$$\lim_{n\to\infty} n\sigma(U^{(\gamma)}, U^{(\delta)}) = m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}.$$

Hence, if $\zeta_1^{(\gamma)} \ne 0$ and $\zeta_1^{(\delta)} \ne 0$, the product moment correlation $\rho(U^{(\gamma)}, U^{(\delta)})$ between $U^{(\gamma)}$ and $U^{(\delta)}$ tends to the limit

$$\lim_{n\to\infty} \rho(U^{(\gamma)}, U^{(\delta)}) = \frac{\zeta_1^{(\gamma,\delta)}}{\sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}}. \tag{6.6}$$

7. Limit theorems for the case of identically distributed X_α's. We shall now study the asymptotic distribution of U-statistics and certain related functions. In this section the vectors X_α will be assumed to be identically distributed. An extension to the case of different parent distributions will be given in section 8.

Following Cramér [2, p. 83] we shall say that a sequence of d.f.'s $F_1(x)$, $F_2(x), \cdots$ converges to a d.f. $F(x)$ if $\lim F_n(x) = F(x)$ in every point at which the one-dimensional marginal limiting d.f.'s are continuous

Let us recall (cf. Cramér [2, p. 312]) that a g-variate normal distribution is called non-singular if the rank r of its covariance matrix is equal to g, and singular if $r < g$.

The following lemma will be used in the proofs

LEMMA 7.1. *Let $V_1, V_2, \cdots$ be an infinite sequence of random vectors $V_n = (V_n^{(1)}, \cdots, V_n^{(g)})$, and suppose that the d.f. $F_n(v)$ of V_n tends to a d.f. $F(v)$ as $n \to \infty$. Let $V_n^{(\gamma)\prime} = V_n^{(\gamma)} + d_n^{(\gamma)}$, where*

$$\lim_{n\to\infty} E\{d_n^{(\gamma)}\}^2 = 0, \qquad (\gamma = 1, \cdots, g). \tag{7.1}$$

Then the d.f. of $V_n' = (V_n^{(1)\prime}, \cdots, V_n^{(g)\prime})$ tends to $F(v)$

This is an immediate consequence of the well-known fact that the d.f. of V_n' tends to $F(v)$ if $d_n^{(\gamma)}$ converges in probability to 0 (cf Cramér [2, p. 299]), since the fulfillment of (7.1) is sufficient for the latter condition

THEOREM 7.1. *Let $X_1, \cdots, X_n$ be n independent, identically distributed random vectors,*

$$X_\alpha = (X_\alpha^{(1)}, \cdots, X_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n).$$

Let

$$\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}), \qquad (\gamma = 1, \cdots, g),$$

be g real-valued functions not involving n, $\Phi^{(\gamma)}$ being symmetric in its $m(\gamma)$ $(\leq n)$ vector arguments $x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)})$, $(\alpha = 1, \cdots, m(\gamma);\ \gamma = 1, \cdots, g)$. Define

$$U^{(\gamma)} = \binom{n}{m(\gamma)}^{-1} \sum{}' \Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}), \qquad (\gamma = 1, \cdots, g), \tag{7.2}$$

where the summation is over all subscripts such that $1 \leq \alpha_1 < \cdots < \alpha_{m(\gamma)} \leq n$ Then, if the expected values

$$\theta^{(\gamma)} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}, \qquad (\gamma = 1, \cdots, g), \tag{7.3}$$

and

$$E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}^2, \qquad (\gamma = 1, \cdots, g), \tag{7.4}$$

exist, the joint d.f. of

$$\sqrt{n}(U^{(1)} - \theta^{(1)}), \cdots, \sqrt{n}(U^{(g)} - \theta^{(g)})$$

tends, as $n \to \infty$, to the g-variate normal d.f. with zero means and covariance matrix $(m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)})$, where $\zeta_1^{(\gamma,\delta)}$ is defined by (6.3). The limiting distribution is non-singular if the determinant $|\zeta_1^{(\gamma,\delta)}|$ is positive.

Before proving Theorem 7.1, a few words may be said about its meaning and its relation to well-known results

For $g = 1$, Theorem 7 1 states that the distribution of a U-statistic tends, under certain conditions, to the normal form. For $m = 1$, U is the sum of n inde-

pendent random variables, and in this case Theorem 7.1 reduces to the Central Limit Theorem for such sums For $m > 1$, U is a sum of random variables which, in general, are not independent Under certain assumptions about the function $\Phi(x_1, \cdots, x_m)$ the asymptotic normality of U can be inferred from the Central Limit Theorem by well-known methods. If, for instance, Φ is a polynomial (as in the case of the k-statistics or the unbiased estimates of moments), U can be expressed as a polynomial in moments about the origin which are sums of independent random variables, and for this case the tendency to normality of U can easily be shown (cf. Cramér [2, p. 365]).

Theorem 7.1 generalizes these results, stating that in the case of independent and identically distributed X_α's the existence of $E\{\Phi^2(X_1, \cdots, X_m)\}$ is sufficient for the asymptotic normality of U. No regularity conditions are imposed on the function Φ. This point is important for some applications (cf. section 9).

Theorem 7.1 and the following theorems of sections 7 and 8 are closely related to recent results of von Mises [18] which were published after this paper was essentially completed. It will be seen below (Theorem 7.4) that the limiting distribution of $\sqrt{n}[U - \theta(F)]$ is the same as that of $\sqrt{n}[\theta(S) - \theta(F)]$ (cf. (4.5)) if the variance of $\theta(S)$ exists. $\theta(S)$ is a differentiable statistical function in the sense of von Mises, and by Theorem I of [18], $\sqrt{n}[\theta(S) - \theta(F)]$ is asymptotically normal if certain conditions are satisfied. It will be found that in certain cases, for instance if the kernel Φ of θ is a polynomial, the conditions of the theorems of sections 7 and 8 are somewhat weaker than those of von Mises' theorem. Though von Mises' paper is concerned with functionals of univariate d.f.'s only, its results can easily be extended to the multivariate case.

For the particular case of a discrete population (where F is a step function), U and $\theta(S)$ are polynomials in the sample frequencies, and their asymptotic distribution may be inferred from the fact that the joint distribution of the frequencies tends to the normal form (cf. also von Mises [18]).

In Theorem 7.1 the functions $\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)})$ are supposed to be symmetric. Since, as has been seen in section 4, any U-statistic with non-symmetric kernel can be written in the form (4.4) with a symmetric kernel, this restriction is not essential and has been made only for the sake of convenience. Moreover, in the condition of the existence of $E\{\Phi^2(X_1, \cdots, X_m)\}$, the symmetric kernel may be replaced by a non-symmetric one. For, if Φ is non-symmetric, and Φ_0 is the symmetric kernel defined by (3.3), $E\{\Phi_0^2(X_1, \cdots, X_m)\}$ is a linear combination of terms of the form $E\{\Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}) \Phi(X_{\beta_1}, \cdots, X_{\beta_m})\}$, whose existence follows from that of $E\{\Phi^2(X_1, \cdots, X_m)\}$ by Schwarz's inequality.

If the regular functional $\theta(F)$ is stationary for $F = F_0$, that is, if $\zeta_1 = \zeta_1(F_0) = 0$ (cf. section 5), the limiting normal distribution of $\sqrt{n}(U - \theta)$ is, according to Theorem 7.1, singular, that is, its variance is zero. As has been seen in section 5, $\sigma^2(U)$ need not be zero in this case, but may be of some order n^{-c}, $(c = 2, 3, \cdots, m)$, and the distribution of $n^{c/2}(U - \theta)$ may tend to a limiting form which is not normal. According to von Mises [18], it is a limiting distribution of type c, $(c = 2, 3, \cdots)$.

According to Theorem 5.2, $\sigma^2(U)$ exceeds its asymptotic value $m^2\zeta_1/n$ for any finite n Hence, if we apply Theorem 7.1 for approximating the distribution of U when n is large but finite, we underestimate the variance of U For many applications this is undesirable, and for such cases the following theorem, which is an immediate consequence of Theorem 7.1, will be more useful

THEOREM 7 2. *Under the conditions of Theorem* 7 1, *and if*

$$\zeta_1^{(\gamma)} > 0, \qquad (\gamma = 1, \cdots, g),$$

the joint d.f of

$$(U^{(1)} - \theta^{(1)})/\sigma(U^{(1)}), \cdots, (U^{(g)} - \theta^{(g)})/\sigma(U^{(g)})$$

tends, as $n \to \infty$, to the g-variate normal d f. with zero means and covariance matrix $(\rho^{(\gamma,\delta)})$, *where*

$$\rho^{(\gamma,\delta)} = \lim_{n\to\infty} \frac{\sigma(U^{(\gamma)}, U^{(\delta)})}{\sigma(U^{(\gamma)})\sigma(U^{(\delta)})} = \frac{\zeta_1^{(\gamma,\delta)}}{\sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}}, \qquad (\gamma, \delta = 1, \cdots, g)$$

PROOF OF THEOREM 7.1. The existence of (7.4) entails that of

$$\zeta_m^{(\gamma)} = E\{\Phi^{(\gamma)}(X_1, \cdots, X_{m(\gamma)})\}^2 - (\theta^{(\gamma)})^2$$

which, by (5.19), (5.20) and (6 6), is sufficient for the existence of

$$\zeta_1^{(\gamma)}, \cdots, \zeta_{m-1}^{(\gamma)}, \quad \text{of} \quad \sigma^2(U^{(\gamma)}), \quad \text{and of} \quad \zeta_1^{(\gamma,\delta)} \leq \sqrt{\zeta_1^{(\gamma)}\zeta_1^{(\delta)}}$$

Now, consider the g quantities

$$Y^{(\gamma)} = \frac{m(\gamma)}{\sqrt{n}} \sum_{\alpha=1}^{n} \Psi_1^{(\gamma)}(X_\alpha), \qquad (\gamma = 1, \cdots, g)$$

where $\Psi_1^{(\gamma)}(x)$ is defined by (6.2). $Y^{(1)}, \cdots, Y^{(g)}$ are sums of n independent, random variables with zero means, whose covariance matrix, by virtue of (6.3), is

$$\{\sigma(Y^{(\gamma)}, Y^{(\delta)})\} = \{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}. \tag{7.5}$$

By the Central Limit Theorem for vectors (cf. Cramér [1, p 112]), the joint d.f. of $(Y^{(1)}, \cdots, Y^{(g)})$ tends to the normal g-variate d.f. with the same means and covariances.

Theorem 7 1 will be proved by showing that the g random variables

$$Z^{(\gamma)} = \sqrt{n}(U^{(\gamma)} - \theta^{(\gamma)}), \qquad (\gamma = 1, \cdots, g), \tag{7.6}$$

have the same joint limiting distribution as $Y^{(1)}, \cdots, Y^{(g)}$

According to Lemma 7.1 it is sufficient to show that

$$\lim_{n\to\infty} E(Z^{(\gamma)} - Y^{(\gamma)})^2 = 0, \qquad (\gamma = 1, \cdots, n). \tag{7.7}$$

For proving (7.7), write

$$E\{Z^{(\gamma)} - Y^{(\gamma)}\}^2 = E\{Z^{(\gamma)}\}^2 + E\{Y^{(\gamma)}\}^2 - 2E\{Z^{(\gamma)}Y^{(\gamma)}\} \tag{7.8}$$

By (5.13) we have

$$E\{Z^{(\gamma)}\}^2 = n\sigma^2(U^{(\gamma)}) = m^2(\gamma)\zeta_1^{(\gamma)} + O(n^{-1}), \tag{7.9}$$

and from (7.5),

$$E\{Y^{(\gamma)}\}^2 = m^2(\gamma)\zeta_1^{(\gamma)}. \tag{7.10}$$

By (7.2) and (6.1) we may write for (7.6)

$$Z^{(\gamma)} = \sqrt{n}\binom{n}{m(\gamma)}^{-1} \Sigma' \Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}}),$$

and hence

$$E\{Z^{(\gamma)}Y^{(\gamma)}\} = m(\gamma)\binom{n}{m(\gamma)}^{-1} \sum_{\alpha=1}^{n} \sum{}' E\{\Psi_1^{(\gamma)}(X_\alpha)\Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})\}.$$

The term

$$E\{\Psi_1^{(\gamma)}(X_\alpha)\Psi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})\}$$

is $= \zeta_1^{(\gamma)}$ if

$$\alpha_1 = \alpha \quad \text{or} \quad \alpha_2 = \alpha \cdots \quad \text{or} \quad \alpha_{m(\gamma)} = \alpha \tag{7.11}$$

and 0 otherwise. For a fixed α, the number of sets $\{\alpha_1, \cdots, \alpha_{m(\gamma)}\}$ such that $1 \leq \alpha_1 < \cdots < \alpha_{m(\gamma)} \leq n$ and (7.11) is satisfied, is $\binom{n-1}{m(\gamma)-1}$. Thus,

$$E\{Z^{(\gamma)}Y^{(\gamma)}\} = m(\gamma)\binom{n}{m(\gamma)}^{-1} n \binom{n-1}{m(\gamma)-1}\zeta_1^{(\gamma)} = m^2(\gamma)\zeta_1^{(\gamma)}. \tag{7.12}$$

On inserting (7.9), (7.10), and (7.12) in (7.8), we see that (7.7) is true.

The concluding remark in Theorem 7.1 is a direct consequence of the definition of a non-singular distribution. The proof of Theorem 7.1 is complete.

Theorems 7.1 and 7.2 deal with the asymptotic distribution of $U^{(1)}, \cdots, U^{(g)}$, which are unbiased estimates of $\theta^{(1)}, \cdots, \theta^{(g)}$. The unbiasedness of a statistic is, of course, irrelevant for its asymptotic behavior, and the application of Lemma 7.1 leads immediately to the following extension of Theorem 7.1 to a larger class of statistics.

THEOREM 7.3. *Let*

$$U^{(g)\prime} = U^{(g)} + \frac{b_n^{(\gamma)}}{\sqrt{n}}, \qquad (\gamma = 1, \cdots, g), \tag{7.13}$$

where $U^{(\gamma)}$ is defined by (7.2) and $b_n^{(\gamma)}$ is a random variable. If the conditions of Theorem 7.1 are satisfied, and $\lim E\{b_n^{(\gamma)}\}^2 = 0$, $(\gamma = 1, \cdots, g)$, *then the joint distribution of*

$$\sqrt{n}(U^{(1)\prime} - \theta^{(1)}), \cdots, \sqrt{n}(U^{(g)\prime} - \theta^{(g)})$$

tends to the normal distribution with zero means and covariance matrix

$$\{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}.$$

This theorem applies, in particular, to the regular functionals $\theta(S)$ of the sample d.f.,

$$\theta(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_m=1}^{n} \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}),$$

in the case that the variance of $\theta(S)$ exists. For we may write

$$n^m\theta(S) = \binom{n}{m} U + \Sigma^* \Phi(X_{\alpha_1}, \cdots, X_{\alpha_m}),$$

where the sum Σ^* is extended over all m-tuplets $(\alpha_1, \cdots, \alpha_m)$ in which at least one equality $\alpha_i = \alpha_j (i \neq j)$ is satisfied. The number of terms in Σ^* is of order n^{m-1}. Hence

$$\theta(S) - U = \frac{1}{n} D,$$

where the expected value $E\{D^2\}$, whose existence follows from that of $\sigma^2\{\theta(S)\}$, is bounded for $n \to \infty$. Thus, if we put $U^{(\gamma)'} = \theta^{(\gamma)}(S)$, the conditions of Theorem 7.3 are fulfilled. We may summarize this result as follows:

THEOREM 7.4. *Let $X_1, \cdots, X_n$ be a random sample from an r-variate population with d.f. $F(x) = F(x^{(1)}, \cdots, x^{(r)})$, and let*

$$\theta^{(\gamma)}(F) = \int \cdots \int \Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)}) \, dF(x_1) \cdots dF(x_{m^{(\gamma)}}), \qquad (\gamma = 1, \cdots, g),$$

be g regular functionals of F, where $\Phi^{(\gamma)}(x_1, \cdots, x_{m(\gamma)})$ is symmetric in the vectors $x_1, \cdots, x_{m(\gamma)}$ and does not involve n. If $S(x)$ is the d.f. of the random sample, and if the variance of

$$\theta^{(\gamma)}(S) = \frac{1}{n^m} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_{m(\gamma)}=1}^{n} \Phi^{(\gamma)}(X_{\alpha_1}, \cdots, X_{\alpha_{m(\gamma)}})$$

exists, the joint d.f. of

$$\sqrt{n}\{\theta^{(1)}(S) - \theta^{(1)}(F)\}, \cdots, \sqrt{n}\{\theta^{(g)}(S) - \theta^{(g)}(F)\}$$

tends to the g-variate normal d.f. with zero means and covariance matrix

$$\{m(\gamma)m(\delta)\zeta_1^{(\gamma,\delta)}\}.$$

The following theorem is concerned with the asymptotic distribution of a function of statistics of the form U or U'.

THEOREM 7.5. *Let $(U') = (U^{(1)'}, \cdots, U^{(g)'})$ be a random vector, where $U^{(\gamma)'}$ is defined by (7.13), and suppose that the conditions of Theorem 7.3 are satisfied. If the function $h(y) = h(y^{(1)}, \cdots, y^{(g)})$ does not involve n and is continuous together with its second order partial derivatives in some neighborhood of the point $(y) = (\theta) = (\theta^{(1)}, \cdots, \theta^{(g)})$, then the distribution of the random variable $\sqrt{n}\{h(U') - h(\theta)\}$ tends to the normal distribution with mean zero and variance*

$$\sum_{\gamma=1}^{g} \sum_{\delta=1}^{g} m(\gamma)m(\delta) \left(\frac{\partial h(y)}{\partial y^{(\gamma)}}\right)_{y=\theta} \left(\frac{\partial h(y)}{\partial y^{(\delta)}}\right)_{y=\theta} \zeta_1^{(\gamma,\delta)}.$$

Theorem 7.5 follows from Theorem 7.3 in exactly the same way as the theorem on the asymptotic distribution of a function of moments follows from the fact of their asymptotic normality; cf Cramér [2, p. 366] We shall therefore omit the proof of Theorem 7 5. Since any moment whose variance exists has the form $U' = \theta(S)$ (cf. section 4 and Theorem 7.4), Theorem 7 5 is a generalization of the theorem on a function of moments.

8. Limit theorems for $U(X_1, \cdots, X_n)$ when the X_α's have different distributions. The limit theorems of the preceding section can be extended to the case when the X_α's have different distributions We shall only prove an extension to this case of Theorem 7.1 (or 7.2), confining ourselves, for the sake of simplicity, to the distribution of a single U-statistic.

The extension of Theorems 7 3 and 7 5 with $g = 1$ to this case is immediate. One has only to replace the reference to Theorem 7.1 by that to the following Theorem 8.1, and θ and ζ_1 by $E\{U\}$ and $\zeta_{1,n}$.

THEOREM 8.1. *Let $X_1, \cdots, X_n$ be n independent random vectors of r components, X_α having the d f $F_\alpha(x) = F_\alpha(x^{(1)}, \cdots, x^{(r)})$. Let $\Phi(x_1, \cdots, x_m)$ be a function symmetric in its m vector arguments $x_\beta = (x_\beta^{(1)}, \cdots, x_\beta^{(r)})$ which does not involve n, and let*

$$\bar{\Psi}_{1(\nu)}(x) = \binom{n-1}{m-1}^{-1} \sum_{(\neq\nu)}' \Psi_{1(\nu)\alpha_1, \cdots, \alpha_{m-1}}(x), \qquad (\nu = 1, \cdots, n), \tag{8.1}$$

where Ψ is defined by (5.15), and the summation is extended over all subscripts α such that

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_{m-1} \leq n, \qquad \alpha_i \neq \nu, \qquad (i = 1, \cdots, m).$$

Suppose that there is a number A such that for every $n = 1, 2, \cdots$

$$\int \cdots \int \Phi^2(x_1, \cdots, x_m)\, dF_{\alpha_1}(x_1) \cdots dF_{\alpha_m}(x_m) < A, \qquad (1 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_m \leq n), \tag{8.2}$$

that

$$E\,|\bar{\Psi}_{1(\nu)}^3(X_\nu)| < \infty, \qquad (\nu = 1, 2, \cdots, n), \tag{8.3}$$

and

$$\lim_{n\to\infty} \sum_{\nu=1}^{n} E\,|\bar{\Psi}_{1(\nu)}^3(X_\nu)| \Big/ \left\{\sum_{\nu=1}^{n} E\{\bar{\Psi}_{1(\nu)}^2(X_\nu)\}\right\}^{3/2} = 0. \tag{8.4}$$

Then, as $n \to \infty$, the d.f. of $(U - E\{U\})/\sigma(U)$ tends to the normal d.f. with mean 0 and variance 1.

The proof is similar to that of Theorem 7.1.

Let

$$W = \frac{m}{n} \sum_{\nu=1}^{n} \bar{\Psi}_{1(\nu)}(X_\nu)$$

It will be shown that

(a) the d.f. of

$$V = \frac{W - E\{W\}}{\sigma(W)}$$

tends to the normal d.f. with mean 0 and variance 1, and that

(b) the d.f. of

$$V' = \frac{U - E\{U\}}{\sigma(U)}$$

tends to the same limit as the d.f. of V.

Part (a) follows immediately from (8.3) and (8.4) by Liapounoff's form of the Central Limit Theorem.

According to Lemma 7.1, (b) will be proved when it is shown that

$$\lim_{n\to\infty} E\{V' - V\}^2 = \lim\left\{2 - 2\,\frac{\sigma(U, W)}{\sigma(U)\sigma(W)}\right\} = 0$$

or

$$\lim_{n\to\infty} \frac{\sigma(U, W)}{\sigma(U)\sigma(W)} = 1. \tag{8.5}$$

Let c be an integer, $1 \leq c \leq m$, and write

$$x = (x_1, \cdots, x_c), \qquad y = (y_1, \cdots, y_{m-c}), \qquad z = (z_1, \cdots, z_{m-c})$$

$$F_{(\alpha)}(x) = F_{\alpha_1}(x_1) \cdots F_{\alpha_c}(x_c), \qquad F_{(\beta)}(y) = F_{\beta_1}(y_1) \cdots F_{\beta_{m-c}}(y_{m-c}),$$

$$F_{(\gamma)}(z) = F_{\gamma_1}(z_1) \cdots F_{\gamma_{m-c}}(z_{m-c})$$

Then, by Schwarz's inequality,

$$\int \cdots \int \Phi(x, y)\Phi(x, z)\, dF_{(\alpha)}(x)\, dF_{(\beta)}(y)\, dF_{(\gamma)}(z)$$

$$\leq \left\{\int \cdots \int \Phi^2(x, y)\, dF_{(\alpha)}(x)\, dF_{(\beta)}(y) \cdot \int \cdots \int \Phi^2(x, z)\, dF_{(\alpha)}(x)\, dF_{(\gamma)}(z)\right\}^{\frac{1}{2}},$$

which, by (8.2), is $< A$ for any set of subscripts.

By the inequality for moments, $\theta_{\alpha_1, \cdots, \alpha_m}$, as defined by (5.14), is also uniformly bounded, and applying these inequalities to (5.16), it follows that there exists a number B such that

$$|\zeta_{c(\alpha_1, \cdots, \alpha_c)\beta_1, \cdots, \beta_{m-c}; \gamma_1, \cdots, \gamma_{m-c}}| < B, \qquad (c = 1, \cdots, m), \tag{8.6}$$

for every set of subscripts satisfying the inequalities

$$\alpha_g \neq \alpha_h, \quad \beta_g \neq \beta_h, \quad \gamma_g \neq \gamma_h \quad \text{if } g \neq h, \quad \alpha_i \neq \beta_j, \quad \alpha_i \neq \gamma_j,$$
$$(i = 1, \cdots, c;\ j = 1, \cdots, m - c).$$

Now, we have

$$E\{W\} = 0$$

and

$$(8.7)\qquad \sigma^2(W) = \frac{m^2}{n^2}\sum_{\nu=1}^{n} E\{\bar{\Psi}^2_{1(\nu)}(X_\nu)\}$$

or, inserting (8 1) and recalling (5.16),

$$(8.8)\qquad \sigma^2(W) = \frac{m^2}{n^2}\binom{n-1}{m-1}^{-2}\sum_{\nu=1}^{n}\sum_{(\neq\nu)}{}'\sum_{(\neq\nu)}{}' \zeta_{1(\nu)\alpha_1,\cdots,\alpha_{m-1};\beta_1,\cdots,\beta_{m-1}},$$

the two sums Σ' being over $\alpha_1 < \cdots < \alpha_{m-1}$, $(\alpha_i \neq \nu)$, and $\beta_1 < \cdots < \beta_{m-1}$, $(\beta_i \neq \nu)$, respectively. By (5.17), the sum of the terms whose subscripts $\nu, \alpha_1, \cdots, \alpha_{m-1}, \beta_1, \cdots, \beta_{m-1}$ are all different is equal to

$$\frac{n(n-1)\cdots(n-2m+2)}{(m-1)!(m-1)!}\zeta_{1,n} = n\binom{n-1}{m-1}\binom{n-m}{m-1}\zeta_{1,n}.$$

The number of the remaining terms is of order n^{2m-2}. Since, by (8.6), they are uniformly bounded, we have

$$(8.9)\qquad \sigma^2(W) = \frac{m^2}{n}\zeta_{1,n} + O(n^{-2}).$$

Similarly, we have from (5.18)

$$\sigma^2(U) = \frac{m^2}{n}\zeta_{1,n} + O(n^{-2}),$$

and hence

$$(8.10)\qquad \sigma(U) = \sigma(W) + O(n^{-1}).$$

The covariance of U and W is

$$(8.11)\quad \sigma(U, W) = \binom{n}{m}^{-1}\frac{m}{n}\sum_{\nu=1}^{n}\sum{}' E\{\bar{\Psi}_{1(\nu)}(X_\nu)\Psi_{m(\alpha_1,\cdots,\alpha_m)}(X_{\alpha_1}, \cdots, X_{\alpha_m})\}.$$

All terms except those in which one of the α's $= \nu$, vanish, and for the remaining ones we have, for fixed $\alpha_1, \cdots, \alpha_m$,

$$\begin{aligned} &E\{\bar{\Psi}_{1(\nu)}(X_\nu)\Psi_{m(\alpha_1,\cdots,\alpha_m)}(X_{\alpha_1}, \cdots, X_{\alpha_m})\} \\ &\qquad = \binom{n-1}{m-1}^{-1}\sum_{(\neq\nu)}{}' E\{\Psi_{1(\nu)\beta_1,\cdots,\beta_{m-1}}(X_\nu)\Psi_{1(\nu)\gamma_1,\cdots,\gamma_{m-1}}(X_\nu)\} \\ &\qquad = \binom{n-1}{m-1}^{-1}\sum_{(\neq\nu)}{}' \zeta_{1(\nu)\beta_1,\cdots,\beta_{m-1};\gamma_1,\cdots,\gamma_{m-1}} \end{aligned}$$

where the summation sign refers to the β's, and $\gamma_1, \cdots, \gamma_{m-1}$ are the α's that are $\neq \nu$. Inserting this in (8.11) and comparing the result with (8.8), we see that

$$(8.12)\qquad \sigma(U, W) = \sigma^2(W).$$

From (8.12) and (8.10) we have

$$\frac{\sigma(U, W)}{\sigma(U)\sigma(W)} = \frac{\sigma(W)}{\sigma(U)} = \frac{n\sigma(W)}{n\sigma(W) + O(1)}.$$

Comparing condition (8.4) with (8.7), we see that we must have $n\sigma(W) \to \infty$ as $n \to \infty$. This shows the truth of (8.5). The proof of Theorem 8.1 is complete.

For some purposes the following corollary of Theorem 8.1 will be useful, where the conditions (8.2), (8.3), and (8.4) are replaced by other conditions which are more restrictive, but easier to apply.

THEOREM 8.2. *Theorem* 8.1 *holds if the conditions* (8.2), (8.3), *and* (8.4) *are replaced by the following.*

There exist two positive numbers C, D such that

$$\int \cdots \int |\Phi^3(x_1, \cdots, x_m)| \, dF_{\alpha_1}(x_1) \cdots dF_{\alpha_m}(x_m) < C \tag{8.13}$$

for $\alpha_i = 1, 2, \cdots, (i = 1, \cdots, m)$, *and*

$$\zeta_{1(\nu)\alpha_1, \cdots, \alpha_{m-1}, \beta_1, \cdots, \beta_{m-1}} > D \tag{8.14}$$

for any subscripts satisfying

$$1 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_{m-1}, \qquad 1 \leq \beta_1 < \beta_2 < \cdots < \beta_{m-1}, \qquad 1 \leq \nu \neq \alpha_i, \beta_i.$$

We have to show that (8.2), (8.3), and (8.4) follow from (8.13) and (8.14).

(8.13) implies (8.2) by the inequality for moments. By a reasoning analogous to that used in the previous proof, applying Holder's inequality instead of Schwarz's inequality, it follows from (8.13) that

$$E\,|\bar{\Psi}^3_{1(\nu)}(X_\nu)| < C'. \tag{8.15}$$

On the other hand, by (8.7), (8.8), and (8.14),

$$\sum_{\nu=1}^{n} E\{\bar{\Psi}^2_{1(\nu)}(X_\nu)\} > nD. \tag{8.16}$$

(8.15) and (8.16) are sufficient for the fulfillment of (8.4).

9. Applications to particular statistics.

(a) *Moments and functions of moments.* It has been seen in section 4 that the k-statistics and the unbiased estimates of moments are U-statistics, while the sample moments are regular functionals of the sample d.f. By Theorems 7.1, 8.1, and 7.4 these statistics are asymptotically normally distributed, and by Theorem 7.5 the same is true for a function of moments, if the respective conditions are satisfied. These results are not new (cf., for example, Cramér [2]).

(b) *Mean difference and coefficient of concentration.* If $Y_1, \cdots, Y_n$ are n independent real-valued random variables, Gini's mean difference (without repetition) is defined by

$$d = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} |Y_\alpha - Y_\beta|.$$

If the Y_α's have the same distribution F, the mean of d is

$$\delta = \iint |y_1 - y_2|\, dF(y_1)\, dF(y_2),$$

and the variance, by (5.13) is

$$\sigma^2(d) = \frac{2}{n(n-1)} \{2\zeta_1(\delta)(n-2) + \zeta_2(\delta)\},$$

where

$$\zeta_1(\delta) = \int \left\{ \int |y_1 - y_2|\, dF(y_2) \right\}^2 dF(y_1) - \delta^2, \tag{9.1}$$

$$\zeta_2(\delta) = \iint (y_1 - y_2)^2\, dF(y_1)\, dF(y_2) - \delta^2 = 2\sigma^2(Y) - \delta^2. \tag{9.2}$$

The notation $\zeta_1(\delta)$, $\zeta_2(\delta)$ serves to indicate the relation of these functionals of F to the functional $\delta(F)$, δ is here merely the symbol of the functional, not a particular value of it In a similar way we shall write $\Phi(y_1, y_2 \mid \delta) = |y_1 - y_2|$, etc. When there is danger of confusing $\zeta_1(\delta)$ with $\zeta_1(F)$, we may write $\zeta_1(F \mid \delta)$

U. S. Nair [19] has evaluated $\sigma^2(d)$ for several particular distributions.

By Theorem 7 1, $\sqrt{n}(d - \delta)$ is asymptotically normal if $\zeta_2(\delta)$ exists.

If $Y_1, \cdots, Y_n$ do not assume negative values, the coefficient of concentration (cf Gini [8]) is defined by

$$G = \frac{d}{2\bar{Y}},$$

where $\bar{Y} = \Sigma Y_\alpha / n$ G is a function of two U-statistics. If the Y_α's are identically distributed, if $E\{Y^2\}$ exists, and if $\mu = E\{Y\} > 0$, then, by Theorem 7.5, $\sqrt{n}(G - \delta/2\mu)$ tends to be normally distributed with mean 0 and variance

$$\frac{\delta^2}{4\mu^4}\zeta_1(\mu) - \frac{\delta}{\mu^3}\zeta_1(\mu, \delta) + \frac{1}{\mu^2}\zeta_1(\delta),$$

where

$$\zeta_1(\mu) = \int y^2\, dF(y) - \mu^2 = \sigma^2(Y),$$

$$\zeta_1(\mu, \delta) = \iint y_1 |y_1 - y_2|\, dF(y_1)\, dF(y_2) - \mu\delta,$$

and $\zeta_1(\delta)$ is given by (9.1).

(c) *Functions of ranks and of the signs of variate differences.* Let $s(u)$ be the signum function,

$$s(u) = \begin{matrix} -1 \text{ if } u < 0, \\ 0 \text{ if } u = 0; \\ 1 \text{ if } u > 0, \end{matrix} \tag{9.3}$$

and let

$$(9.4)\qquad c(u) = \tfrac{1}{2}\{1 + s(u)\} = \begin{cases} 0 \text{ if } u < 0, \\ \tfrac{1}{2} \text{ if } u = 0, \\ 1 \text{ if } u > 0. \end{cases}$$

If

$$x_\alpha = (x_\alpha^{(1)}, \cdots, x_\alpha^{(r)}), \qquad (\alpha = 1, \cdots, n)$$

is a sample of n vectors of r components, we may define the *rank* $R_\alpha^{(i)}$ of $x_\alpha^{(i)}$ by

$$(9.5)\qquad \begin{aligned} R_\alpha^{(i)} &= \tfrac{1}{2} + \sum_{\beta=1}^{n} c(x_\alpha^{(i)} - x_\beta^{(i)}) \\ &= \frac{n+1}{2} + \tfrac{1}{2}\sum_{\beta=1}^{n} s(x_\alpha^{(i)} - x_\beta^{(i)}), \quad (i = 1, \cdots, r) \end{aligned}$$

If the numbers $x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)}$ are all different, the smallest of them has rank 1, the next smallest rank 2, etc. If some of them are equal, the rank as defined by (9.5) is known as the mid-rank.

Any function of the ranks is a function of expressions $c(x_\alpha^{(i)} - x_\beta^{(i)})$ or $s(x_\alpha^{(i)} - x_\beta^{(i)})$.

Conversely, since

$$s(x_\alpha^{(i)} - x_\beta^{(i)}) = s(R_\alpha^{(i)} - R_\beta^{(i)}),$$

any function of expressions $s(x_\alpha^{(i)} - x_\beta^{(i)})$ or $c(x_\alpha^{(i)} - x_\beta^{(i)})$ is a function of the ranks.

Consider a regular functional $\theta(F)$ whose kernel $\Phi(x_1, \cdots, x_m)$ depends only on the signs of the variate differences,

$$(9.6)\qquad s(x_\alpha^{(i)} - x_\beta^{(i)}), \qquad (\alpha, \beta = 1, \cdots, m, i = 1, \cdots, r).$$

The corresponding U-statistic is a function of the ranks of the sample variates.

The function Φ can take only a finite number of values, $c_1, \cdots, c_N$, say. If $\pi_i = P\{\Phi = c_i\}$, $(i = 1, \cdots, N)$, we have

$$\theta = c_1\pi_1 + \cdots + c_N\pi_N, \qquad \sum_{i=1}^{N} \pi_i = 1.$$

π_i is a regular functional whose kernel $\Phi_i(x_1, \cdots, x_m)$ is equal to 1 or 0 according to whether $\Phi = c_i$ or $\neq c_i$. We have

$$\Phi = c_1\Phi_1 + \cdots + c_N\Phi_N .$$

In order that $\theta(F)$ exist, the c_i must be finite, and hence Φ is bounded. Therefore, $E\{\Phi^2\}$ exists, and if $X_1, X_2, \cdots$ are identically distributed, the d.f. of $\sqrt{n}(U - \theta)$ tends, by Theorem 7.1, to a normal d.f. which is non-singular if $\zeta_1 > 0$.

In the following we shall consider several examples of such functionals.

(d) *Difference sign correlation.* Consider the bivariate sample

$$(9.7) \qquad (x_1^{(1)}, x_1^{(2)}), (x_2^{(1)}, x_2^{(2)}), \cdots, (x_n^{(1)}, x_n^{(2)}).$$

To each two members of this sample corresponds a pair of signs of the differences of the respective variables,

$$(9.8) \qquad s(x_\alpha^{(1)} - x_\beta^{(1)}), s(x_\alpha^{(2)} - x_\beta^{(2)}), \qquad (\alpha \neq \beta; \alpha, \beta = 1, \cdots, n).$$

(9.8) is a population of $n(n-1)$ pairs of difference signs. Since

$$\sum_{\alpha \neq \beta} s(x_\alpha^{(i)} - x_\beta^{(i)}) = 0, \qquad (i = 1, 2),$$

the covariance t of the difference signs (9.8) is

$$(9.9) \qquad t = \frac{1}{n(n-1)} \sum_{\alpha \neq \beta} s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\beta^{(2)})$$

t will be briefly referred to as the *difference sign covariance* of the sample (9.7).

If all $x^{(1)}$'s and all $x^{(2)}$'s are different, we have

$$\sum_{\alpha \neq \beta} s^2(x_\alpha^{(i)} - x_\beta^{(i)}) = n(n-1), \qquad (i = 1, 2),$$

and then t is the product moment correlation of the difference signs.

It is easily seen that t is a linear function of the number of inversions in the permutation of the ranks of $x^{(1)}$ and $x^{(2)}$.

The statistic t has been considered by Esscher [6], Lindeberg [15], [16], Kendall [12], and others.

t is a U-statistic. As a function of a random sample from a bivariate population, t is an unbiased estimate of the regular functional of degree 2,

$$(9.10) \qquad \tau = \iiiint s(x_1^{(1)} - x_2^{(1)}) s(x_1^{(2)} - x_2^{(2)})\, dF(x_1)\, dF(x_2).$$

τ is the covariance of the signs of differences of the corresponding components of $X_1 = (X_1^{(1)}, X_1^{(2)})$ and $X_2 = (X_2^{(1)}, X_2^{(2)})$ in the population of pairs of independent vectors X_1, X_2 with identical d.f. $F(x) = F(x^{(1)}, x^{(2)})$. If $F(x^{(1)}, x^{(2)})$ is continuous, τ is the product moment correlation of the difference signs

Two points (or vectors), $(x_1^{(1)}, x_1^{(2)})$ and $(x_2^{(1)}, x_2^{(2)})$ are called concordant or discordant according to whether

$$(x_1^{(1)} - x_2^{(1)})(x_1^{(2)} - x_2^{(2)})$$

is positive or negative. If $\pi^{(c)}$ and $\pi^{(d)}$ are the probabilities that a pair of vectors drawn at random from the population is concordant or discordant, respectively, we have from (9.10)

$$\tau = \pi^{(c)} - \pi^{(d)}.$$

If $F(x^{(1)}, x^{(2)})$ is continuous, we have $\pi^{(c)} + \pi^{(d)} = 1$, and hence

$$(9.11) \qquad \tau = 2\pi^{(c)} - 1 = 1 - 2\pi^{(d)}.$$

If we put

$$\bar{F}(x^{(1)}, x^{(2)}) = \tfrac{1}{4}\{F(x^{(1)} - 0, x^{(2)} - 0) + F(x^{(1)} - 0, x^{(2)} + 0) + F(x^{(1)} + 0, x^{(2)} - 0) + F(x^{(1)} + 0, x^{(2)} + 0)\}, \tag{9.12}$$

we have

$$\Phi_1(x \mid \tau) = 1 - 2\bar{F}(x^{(1)}, \infty) - 2\bar{F}(\infty, x^{(2)}) + 4\bar{F}(x^{(1)}, x^{(2)}), \tag{9.13}$$

and we may write

$$\tau = E\{\Phi_1(X_1 \mid \tau)\}. \tag{9.14}$$

The variance of t is, by (5.13),

$$\sigma^2(t) = \frac{2}{n(n-1)}\{2\zeta_1(\tau)(n-2) + \zeta_2(\tau)\}, \tag{9.15}$$

where

$$\zeta_1(\tau) = E\{\Phi_1^2(X_1 \mid \tau)\} - \tau^2, \tag{9.16}$$

$$\zeta_2(\tau) = E\{s^2(X_1^{(1)} - X_2^{(1)})s^2(X_1^{(2)} - X_2^{(2)})\} - \tau^2. \tag{9.17}$$

If $F(x^{(1)}, x^{(2)})$ is continuous, we have $\zeta_2(\tau) = 1 - \tau^2$, and $\bar{F}(x^{(1)}, x^{(2)})$ in (9.13) may be replaced by $F(x^{(1)}, x^{(2)})$.

The variance of a linear function of t has been given for the continuous case by Lindeberg [15], [16].

If $X^{(1)}$ and $X^{(2)}$ are independent and have a continuous d f , we find $\zeta_1(\tau) = \frac{1}{9}$, $\zeta_2(\tau) = 1$, and hence

$$\sigma^2(t) = \frac{2(2n+5)}{9n(n-1)}. \tag{9.18}$$

In this case the distribution of t is independent of the univariate distributions of $X^{(1)}$ and $X^{(2)}$. This is, however, no longer true if the independent variables are discontinuous Then it appears that $\sigma^2(t)$ depends on $P\{X_1^{(i)} = X_2^{(i)}\}$ and $P\{X_1^{(i)} = X_2^{(i)} = X_3^{(i)}\}$, $(i = 1, 2)$

By Theorem 7 1, the d.f. of $\sqrt{n}(t - \tau)$ tends to the normal form This result has first been obtained for the particular case that all permutations of the ranks of $X^{(1)}$ and $X^{(2)}$ are equally probable, which corresponds to the independence of the continuous random variables $X^{(1)}$, $X^{(2)}$ (Kendall [12]). In this case t can be represented as a sum of independent random variables (cf. Dantzig [5] and Feller [7]). In the general case the asymptotic normality of t has been shown by Daniels and Kendall [4] and the author [10].

The functional $\tau(F)$ is stationary (and hence the normal limiting distribution of $\sqrt{n}(t - \tau)$ singular) if $\zeta_1 = 0$, which, in the case of a continuous F, means that the equation $\Phi_1(X \mid \tau) = \tau$ or

$$4F(X^{(1)}, X^{(2)}) = 2F(X^{(1)}, \infty) + 2F(\infty, X^{(2)}) - 1 + \tau \tag{9.19}$$

is satisfied with probability 1. This is the case if $X^{(2)}$ is an increasing function of $X^{(1)}$. Then $t = \tau = 1$ with probability 1, and $\sigma^2(t) = 0$ A case where (9.19) is fulfilled and $\sigma^2(t) > 0$ is the following: $X^{(1)}$ is uniformly distributed in the interval (0, 1), and

$$(9.20) \qquad X^{(2)} = X^{(1)} + \tfrac{1}{2} \text{ if } 0 \le X^{(1)} < \tfrac{1}{2}, \; X^{(2)} = X^{(1)} - \tfrac{1}{2} \text{ if } \tfrac{1}{2} \le X^{(1)} \le 1$$

In this case $\tau = 0$, $\zeta_2 = 1$, $\sigma^2(t) = 2/n(n-1)$.

(e) *Rank correlation and grade correlation* If in the sample $\{(x_\alpha^{(1)}, x_\alpha^{(2)})\}$, $(\alpha = 1, \cdots, n)$, all $x_\alpha^{(1)}$'s and all $x_\alpha^{(2)}$'s are different, the rank correlation coefficient, which we denote by k', is given by

$$k' = \frac{12}{n^3 - n} \sum_{\alpha=1}^{n} \left(R_\alpha^{(1)} - \frac{n+1}{2}\right)\left(R_\alpha^{(2)} - \frac{n+1}{2}\right).$$

Inserting (9.5) we have

$$k' = \frac{3}{n^3 - n} \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} \sum_{\gamma=1}^{n} s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\gamma^{(2)})$$

or

$$(9.21) \qquad k' = \frac{(n-2)k + 3t}{n+1}$$

where t is the difference sign covariance (9.9), and

$$k = \frac{3}{n(n-1)(n-2)} \sum{}'' s(x_\alpha^{(1)} - x_\beta^{(1)}) s(x_\alpha^{(2)} - x_\gamma^{(2)}),$$

the summation being over all different subscripts α, β, γ.

k is a U-statistic, and as a function of a random sample from a population with d.f. F, k is an unbiased estimate of the regular functional of degree 3,

$$(9.22) \qquad \begin{aligned} \kappa &= 3 \int \cdots \int s(x_1^{(1)} - x_2^{(1)}) s(x_1^{(2)} - x_3^{(2)}) \, dF(x_1) \, dF(x_2) \, dF(x_3) \\ &= 3 \iint \{2\bar{F}^{(1)}(x^{(1)}) - 1\}\{2\bar{F}^{(2)}(x^{(2)}) - 1\} \, dF(x), \end{aligned}$$

where $\bar{F}^{(1)}(x^{(1)}) = \bar{F}(x^{(1)}, \infty)$, $\bar{F}^{(2)}(x^{(2)}) = \bar{F}(\infty, x^{(2)})$.

If F is continuous, we have

$$\int \bar{F}^{(i)}(y) \, d\bar{F}^{(i)}(y) = \int_0^1 u \, du = \tfrac{1}{2},$$

$$\int \{\bar{F}^{(i)}(y) - \tfrac{1}{2}\}^2 \, d\bar{F}^{(i)}(u) = \int_0^1 (u - \tfrac{1}{2})^2 \, du = \tfrac{1}{12}, \qquad (i = 1, 2),$$

and in this case κ is the coefficient of correlation between the random variables

$$U^{(1)} = F^{(1)}(X^{(1)}), \qquad U^{(2)} = F^{(2)}(X^{(2)}).$$

$U^{(i)}$ has been termed the grade of the continuous variable $X^{(i)}$, and in the general case $\bar{F}^{(i)}(X^{(i)})$ may be called the grade of $X^{(i)}$ (cf., for instance, G. U. Yule and M. G. Kendall [22, p. 150]). In general, κ is 12 times the covariance of the grades

From (9.21) we have for the expected value of k',

$$E\{k'\} = \frac{(n-2)\kappa + 3\tau}{n+1}.$$

In the continuous case the rank correlation coefficient k' is an estimate of the grade correlation κ, which is biased for finite n but unbiased in the limit.

The kernel $3s(x_1^{(1)} - x_2^{(1)})s(x_1^{(2)} - x_3^{(2)})$ of κ is not symmetric Denoting by $\Phi(x_1, x_2, x_3 \mid \kappa)$ the symmetric kernel of κ, we have

$$\Phi(x_1, x_2, x_3 \mid \kappa) = \frac{1}{2} \sum_{\substack{\alpha \neq \beta \neq \gamma \\ \alpha \neq \gamma}}^{1,2,3} s(x_\alpha^{(1)} - x_\beta^{(1)})s(x_\alpha^{(2)} - x_\gamma^{(2)}) \tag{9.23}$$

For computing κ and the constants ζ, an alternative expression for κ and Φ is sometimes more convenient. From three two-dimensional vectors x_1, x_2, x_3 we can form three pairs (x_1, x_2), (x_1, x_3), and (x_2, x_3). The number of concordant pairs among them can be 3, 2, 1, or 0 If γ is the probability that among the three pairs formed from three random elements of the population at least 2 are concordant, we have, if the d f. F is continuous,

$$\kappa = 2\gamma - 1. \tag{9.24}$$

This is analogous to the expression (9.11) for τ.

The truth of (9 24) can be seen as follows: From the definition of γ we have

$$\gamma = E\{\Phi(x_1, x_2, x_3 \mid \gamma)\},$$

where $\Phi(x_1, x_2, x_3 \mid \gamma)$ is $= 1$ if at least two of the three expressions

$$(x_\alpha^{(1)} - x_\beta^{(1)})(x_\alpha^{(2)} - x_\beta^{(2)}), \qquad (\alpha < \beta;\ \alpha, \beta = 1, 2, 3) \tag{9.25}$$

are positive, and equal to zero, if no more than one of them is positive Since, by the continuity of F, we may neglect the case of (9 25) being zero, we may write

$$\Phi(x_1, x_2, x_3 \mid \gamma) = c_{12,12}c_{23,23}c_{31,31} + c_{12,12}c_{23,23}c_{31,13} + c_{12,12}c_{23,32}c_{31,31} + c_{12,21}c_{23,23}c_{31,31},$$

where

$$c_{\alpha,\beta,\gamma,\delta} = c[(x_\alpha^{(1)} - x_\beta^{(1)})(x_\gamma^{(2)} - x_\delta^{(2)})]$$

and $c(u)$ is defined by (9.4).

$\Phi(x_1, x_2, x_3 \mid \gamma)$ is symmetric in x_1, x_2, x_3.

The identity

$$\Phi(x_1, x_2, x_3 \mid \kappa) = 2\Phi(x_1, x_2, x_3 \mid \gamma) - 1 \tag{9.26}$$

can be shown to hold either by algebraical calculation using (9 4) or by direct computation of each side for the different positions of the three points x_1, x_2, x_3.

From (9.26) it appears that in the continuous case the symmetric kernel $\Phi(x_1, x_2, x_3 \mid \kappa)$ can assume only two values, -1 and $+1$.

The variance of k is, according to (5.13),

$$\sigma^2(k) = \frac{6}{n(n-1)(n-2)} \left\{ 3 \binom{n-3}{2} \zeta_1(\kappa) + 3(n-3)\zeta_2(\kappa) + \zeta_3(\kappa) \right\},$$

where

$$\zeta_1(\kappa) = E\{\Phi_1^2(X_1 \mid \kappa)\} - \kappa^2,$$

$$\zeta_2(\kappa) = E\{\Phi_2^2(X_1, X_2 \mid \kappa)\} - \kappa^2,$$

$$\zeta_3(\kappa) = E\{\Phi^2(X_1, X_2, X_3 \mid \kappa)\} - \kappa^2,$$

$$\Phi_1(x_1 \mid \kappa) = E\{\Phi(x_1, X_2, X_3 \mid \kappa)\},$$

$$\Phi_2(x_1, x_2 \mid \kappa) = E\{\Phi(x_1, x_2, X_3 \mid \kappa)\}.$$

We find for the continuous case

$$\zeta_3(\kappa) = 1 - \kappa^2,$$

$$\begin{aligned} \text{(9.27)} \quad \Phi_1(x_1 \mid \kappa) = {} & [1 - 2F(x_1^{(1)}, \infty)][1 - 2F(\infty, x_1^{(2)})] - 2F(x_1^{(1)}, \infty) \\ & - 2F(\infty, x_1^{(2)}) + 4\int F(x_1^{(1)}, y^{(2)})dF(\infty, y^{(2)}) \\ & + 4\int F(y^{(1)}, x_1^{(2)})dF(y^{(1)}, \infty), \end{aligned}$$

$$\begin{aligned} \Phi_2(x_1, x_2 \mid \kappa) = {} & 1 + 2F(x_1^{(1)}, x_2^{(2)}) + 2F(x_2^{(1)}, x_1^{(2)}) - 2c(x_2^{(2)} - x_1^{(2)})F(x_1^{(1)}, \infty) \\ & -2c(x_1^{(2)} - x_2^{(2)})F(x_2^{(1)}, \infty) - 2c(x_2^{(1)} - x_1^{(1)})F(\infty, x_1^{(2)}) \\ & - 2c(x_1^{(1)} - x_2^{(1)})F(\infty, x_2^{(2)}). \end{aligned}$$

If $X^{(1)}, X^{(2)}$ are continuous and independent, we obtain $\kappa = 0$, $\zeta_1 = \frac{1}{9}$, $\zeta_2 = \frac{7}{18}$, $\zeta_3 = 1$, and hence

$$\text{(9.28)} \qquad \sigma^2(k) = \frac{n^2 - 3}{n(n-1)(n-2)}.$$

In the discontinuous case of independence the distribution of k, as that of t, depends on the distributions of $X^{(1)}$ and $X^{(2)}$, and $\sigma^2(k)$ can again be expressed in terms of $P\{X_1^{(i)} = X_2^{(i)}\}$ and $P\{X_1^{(i)} = X_2^{(i)} = X_3^{(i)}\}$, $(i = 1, 2)$.

The variance of the rank correlation coefficient k' is, by (9 21),

$$\text{(9.29)} \qquad \sigma^2(k') = \frac{(n-2)^2 \sigma^2(k) + 6(n-2)\sigma(t, k) + 9\sigma^2(t)}{(n+1)^2}.$$

For $\sigma(t, k)$ we have, according to (6.5),

$$\sigma(t, k) = \frac{6}{n(n-1)} \{(n-3)\zeta_1(\tau, \kappa) + \zeta_2(\tau, \kappa)\},$$

where

$$\zeta_1(\tau, \kappa) = E\{\Phi_1(X_1 \mid \tau)\Phi_1(X_1 \mid \kappa)\} - \tau\kappa,$$

$$\zeta_2(\tau, \kappa) = E\{\Phi(X_1, X_2 \mid \tau)\Phi_2(X_1, X_2 \mid \kappa)\} - \tau\kappa.$$

In the case of independence we see from (9.13) and (9.27) that

$$\Phi_1(x \mid \tau) = \Phi_1(x \mid \kappa) = [1 - 2F(x^{(1)}, \infty)][1 - 2F(\infty, x^{(2)})],$$

and we obtain

(9.30) $$\zeta_1(\tau, \kappa) = \zeta_1(\kappa) = \zeta_1(\tau) = \tfrac{1}{9},$$

$$\zeta_2(\tau, \kappa) = \tfrac{5}{9},$$

(9.31) $$\sigma(t, k) = \frac{2(n+2)}{3n(n-1)}.$$

On inserting (9.28), (9.31) and (9 18) in (9.29), we find

$$\sigma^2(k') = \frac{1}{n-1},$$

in accordance with the result obtained for this case by Student and published by K. Pearson [20].

According to Theorem 7.1, $\sqrt{n}(k - \kappa)$ tends to be normally distributed with mean 0 and variance $9\zeta_1(\kappa)$. The same is true for the distribution of the rank correlation coefficient, k', as follows from Theorem 7.3 in conjunction with (9 21). For the special case of independence the asymptotic normality of k' has been proved by Hotelling and Pabst [11].

From Theorem 7 3 it also follows that the joint distribution of $\sqrt{n}(t - \tau)$ and $\sqrt{n}(k - \kappa)$ (or $\sqrt{n}(k' - \kappa)$) tends to the normal form with the variances $4\zeta_1(\tau)$ and $9\zeta_1(\kappa)$ and the covariance $6\zeta_1(\kappa, \tau)$. In the case of independence we see from (9.30) that the correlation $\rho(t, k)$ between t and k tends to 1, and we have the asymptotic functional relation $3t = 2k$. This result has been conjectured by Kendall and others [14], and proved by Daniels [3]. In general, however, $\rho(t, k)$ does not approach unity. Thus, if $X^{(1)}$ is uniformly distributed in (0, 1), and

(9 32) $$\begin{aligned} X^{(2)} &= \tfrac{1}{2} - X^{(1)} && \text{if } 0 \le X^{(1)} < \tfrac{1}{4}, \\ X^{(2)} &= \tfrac{1}{2} + X^{(1)} && \text{if } \tfrac{1}{4} \le X^{(1)} < \tfrac{1}{2}, \\ X^{(2)} &= X^{(1)} - \tfrac{1}{2} && \text{if } \tfrac{1}{2} \le X^{(1)} < \tfrac{3}{4}, \\ X^{(2)} &= \tfrac{3}{2} - X^{(1)} && \text{if } \tfrac{3}{4} \le X^{(1)} \le 1, \end{aligned}$$

we have $\tau = \kappa = 0$, $\zeta_1(\tau) = 0$, $\zeta_2(\tau) = 1$, $\zeta_1(\kappa) = \frac{1}{16}$, $\zeta_1(\kappa, \tau) = 0$, and hence $\rho(t, k) \to 0$.

(f) *Non-parametric tests of independence* Suppose that the random variables $X^{(1)}$, $X^{(2)}$ have a continuous joint d.f $F(x^{(1)}, x^{(2)})$, and we want to test the hypothesis H_0 that $X^{(1)}$ and $X^{(2)}$ are independent, that is, that

$$F(x^{(1)}, x^{(2)}) = F(x^{(1)}, \infty)\, F(\infty, x^{(2)}).$$

The distribution of any statistic involving only the ranks of the variables does not depend on the d.f. of the population when H_0 is true. For this reason several rank order statistics, among them the difference sign correlation t and the rank correlation k', have been suggested for testing independence

From the preceding results we can obtain the asymptotic power functions of the tests of independence based on t and k'. If H_0 is true, we have $E\{t\} = \tau = 0$, and the critical region of size ϵ of the t-test may be defined by $|t| > c_n$, where c_n is the smallest number satisfying the inequality

$$P\{\, |t| > c_n \mid H_0\} \leq \epsilon. \tag{9.33}$$

By Theorem 7.2 and (9.18) we may write $c_n = 2\lambda_n/3\sqrt{n}$, where λ_n tends to a positive constant λ depending on ϵ.

Since $\sigma^2(t) = O(n^{-1})$, the power function

$$P_n(H) = P\{\, |t| \geq 2\lambda_n/3\sqrt{n} \mid H\}$$

tends to one as $n \to \infty$ for any alternative hypothesis H with $\tau(F) \neq 0$. If, however, $\tau = 0$, we have $\lim P_n(H) < 1$. If $\tau = 0$ and $\zeta_1(\tau) < \frac{1}{9}$, we have even $\lim P_n(H) < \epsilon$, and with respect to these alternatives the test is biased in the limit. Thus, in the case of the distribution (9.20) we have even $P_n(H) \to 0$. In this case there is a functional relationship between the variables, and the distribution must be considered as considerably different from the case of independence.

For the rank correlation test we have a similar result. If c_n' is the smallest number satisfying $P\{\, |k'| > c_n' \mid \mathrm{H}_0\} \leq \epsilon$, we have $c_n' = \lambda_n'/\sqrt{n}$, where $\lim \lambda_n' = \lambda$, and the test is biased in the limit if $\kappa = 0$ and $\zeta_1(\kappa) < \frac{1}{9}$. This is fulfilled in the case of the distribution (9.32), where $\zeta_1(\kappa) = \frac{1}{18}$.

The question arises whether there exist non-parametric tests of independence which are unbiased or unbiased in the limit. This point will be discussed in a separate paper on tests of independence

(g) *Mann's test against trend* Let $Y_1, \cdots, Y_n$ be n independent real-valued random variables, Y_α having the continuous d f. $F_\alpha(y)$, $(\alpha = 1, \cdots, n)$. The hypothesis of randomness,

$$H_1: \quad F_1(y) = \cdots = F_n(y)$$

is to be tested against the alternative hypothesis of a "downward trend,"

$$H_2: \quad F_1(y) < F_2(y) < \cdots < F_n(y).$$

H. B. Mann [17] has suggested a test of H_1 against H_2 based on the number T of inequalities $Y_\alpha < Y_\beta$, where $\alpha < \beta$ We may write

$$2T - \frac{n(n-1)}{2} = \sum_{\alpha<\beta} s(Y_\beta - Y_\alpha) = \sum_{\alpha<\beta} s(\alpha - \beta)s(Y_\alpha - Y_\beta).$$

The U-statistic

$$t = \{4T/n(n-1)\} - 1$$

is the same as (9.9) for the special case when one component is not a random variable

Let

$$\tau_{\alpha\beta} = s(\alpha - \beta) \iint s(y_1 - y_2)\, dF_\alpha(y_1)\, dF_\beta(y_2)$$

$$= s(\alpha - \beta) \left\{ 2 \int F_\beta(y)\, dF_\alpha(y) - 1 \right\}.$$

We have $\tau_{\alpha\beta} = 0$ if H_1 is true and $\tau_{\alpha\beta} < 0$ if H_2 is true.

Since

$$E\{t\} = \tau_n = \frac{2}{n(n-1)} \sum_{\alpha<\beta} \tau_{\alpha\beta},$$

it follows that $E\{t\} = 0$ under H_1 and $E\{t\} < 0$ under H_2.

Mann's test against trend has the power function $P_n(H) = P\{t < a_n \mid H\}$, where a_n is the largest number satisfying $P\{t < a_n \mid H_1\} \leq \epsilon$.

Since $a_n \to 0$ and, by (5.18), $\sigma^2(t) = O(n^{-1})$, it follows from Tchebycheff's inequality that the test is consistent (that is, $P_n(H_2) \to 1$) and hence unbiased in the limit. This has been shown by Mann who also gave sufficient conditions under which the test is unbiased for finite n.

By Theorems 8 1 and 8 2 the distribution of $(t - \tau_n)/\sigma(t)$ is asymptotically normal if certain conditions are satisfied Since (8.2), (8.3) and (8.13) are fulfilled, either of the conditions (8.4) and (8.14) is sufficient

(h) *The coefficient of partial difference sign correlation.* Consider a three-variate sample $x_1, \cdots, x_n$; $x_\alpha = (x_\alpha^{(1)}, x_\alpha^{(2)}, x_\alpha^{(3)})$, $(\alpha = 1, \cdots, n)$. In a similar way as in section 9d we may form the set of the $n(n-1)$ triplets of difference signs,

$$(9.34) \qquad s(x_\alpha^{(1)} - x_\beta^{(1)}), \qquad s(x_\alpha^{(2)} - x_\beta^{(2)}), \qquad s(x_\alpha^{(3)} - x_\beta^{(3)}),$$

$$(\alpha \neq \beta;\ \alpha, \beta = 1, \cdots, n).$$

We shall assume that all $x^{(1)}$'s, $x^{(2)}$'s, and $x^{(3)}$'s are different. Then the triplets (9.34) contain only two different numbers, $+1$ and -1. Hence the regression functions of the three-variate population (9 34) are linear.

If t_{12}, t_{13}, and t_{23} are the difference sign correlations of $\{s(x_\alpha^{(1)} - x_\beta^{(1)}), s(x_\alpha^{(2)} - x_\beta^{(2)})\}$, $\{s(x_\alpha^{(1)} - x_\beta^{(1)}), s(x_\alpha^{(3)} - x_\beta^{(3)})\}$ and $\{s(x_\alpha^{(2)} - x_\beta^{(2)}), s(x_\alpha^{(3)} - x_\beta^{(3)})\}$ respectively, we have for the coefficient $t_{12\cdot3}$ of partial correlation between $s(x_\alpha^{(1)} - x_\beta^{(1)})$ and $s(x_\alpha^{(2)} - x_\beta^{(2)})$ with respect to $s(x_\alpha^{(3)} - x_\beta^{(3)})$,

$$(9.35) \qquad t_{12\cdot3} = \frac{t_{12} - t_{13}\, t_{23}}{\sqrt{(1 - t_{13}^2)(1 - t_{23}^2)}}.$$

This measure of partial correlation has been suggested by Kendall [13] who gave an alternative definition of $t_{12.3}$.

If we have two independent three-dimensional random vectors $X_1 = (X_1^{(1)}, X_1^{(2)}, X_1^{(3)})$ and $X_2 = (X_2^{(1)}, X_2^{(2)}, X_2^{(3)})$ with the same continuous d.f. $F(x^{(1)}, x^{(2)}, x^{(3)})$, the distribution of the difference signs $s(X_1^{(i)} - X_2^{(i)})$, $(i = 1, 2, 3)$, has again linear regression functions, and we may define the partial difference sign correlation

$$\tau_{12.3} = \frac{\tau_{12} - \tau_{13}\tau_{23}}{\sqrt{(1 - \tau_{13}^2)(1 - \tau_{23}^2)}},$$

where τ_{ij} is the difference sign correlation of $X^{(i)}, X^{(j)}$.

If $t_{12.3}$ is a function of a random sample, and if $\tau_{13}^2 \neq 1$, $\tau_{23}^2 \neq 1$, the d.f. of $\sqrt{n}(t_{12.3} - \tau_{12.3})$ tends, by Theorem 7.5, so the normal d.f. with mean zero and variance

$$\sigma_{12.3}^2 = \frac{4}{(1 - \tau_{13}^2)(1 - \tau_{23}^2)} \left\{ \zeta_1(\tau_{12}) + \frac{(\tau_{23} - \tau_{12}\tau_{13})^2}{(1 - \tau_{13}^2)^2} \zeta_1(\tau_{13}) \right.$$
$$+ \frac{(\tau_{13} - \tau_{12}\tau_{23})^2}{(1 - \tau_{23}^2)^2} \zeta_1(\tau_{23}) - 2\frac{\tau_{23} - \tau_{12}\tau_{13}}{1 - \tau_{13}^2} \zeta_1(\tau_{12}, \tau_{13}) - 2\frac{\tau_{13} - \tau_{12}\tau_{23}}{1 - \tau_{23}^2} \zeta_1(\tau_{12}, \tau_{23})$$
$$\left. + 2\frac{(\tau_{23} - \tau_{12}\tau_{13})(\tau_{13} - \tau_{12}\tau_{23})}{(1 - \tau_{13}^2)(1 - \tau_{23}^2)} \zeta_1(\tau_{13}, \tau_{23}) \right\},$$

where

$$\zeta(\tau_{ij}) = E\{\Phi_1^2(X \mid \tau_{ij})\} - \tau_{ij}^2,$$
$$\zeta_1(\tau_{ij}, \tau_{gh}) = E\{\Phi_1(X \mid \tau_{ij})\Phi_1(X \mid \tau_{gh})\} - \tau_{ij}\tau_{gh},$$

and, for instance (cf (9.13)),

$$\Phi_1(x \mid \tau_{12}) = 1 - 2F(x^{(1)}, \infty, \infty) - 2F(\infty, x^{(2)}, \infty) + 4F(x^{(1)}, x^{(2)}, \infty).$$

If $\tau_{13} = \tau_{23} = 0$, we have

$$\sigma_{12.3}^2 = 4\zeta_1(\tau_{12}),$$

and $\sqrt{n}(t_{12.3} - \tau_{12.3})$ has the same limiting distribution as $\sqrt{n}(t_{12} - \tau_{12})$. This is in particular the case when $X^{(1)}, X^{(2)}, X^{(3)}$ are independent

REFERENCES

[1] H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Math, Cambridge, 1937.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946

[3] H. E. Daniels, "The relation between measures of correlation in the universe of sample permutations," *Biometrika*, Vol 33 (1944), pp 129–135

[4] H. E Daniels and M. G. Kendall, "The significance of rank correlations where parental correlation exists," *Biometrika*, Vol 34 (1947), pp 197–208.

[5] G B. Dantzig, "On a class of distributions that approach the normal distribution function," *Annals of Math Stat.*, Vol. 10 (1939) pp 247–253

[6] F. Esscher, "On a method of determining correlation from the ranks of the variates," *Skandinavisk Aktuar tids*, Vol 7 (1924), pp 201–219

[7] W. Feller, "The fundamental limit theorems in probability," *Am Math Soc Bull*, Vol. 51 (1945), pp 800–832.

[8] C Gini, "Sulla misura della concentrazione e della variabilità dei caratteri," *Atti del R Istituto Veneto di S.L.A.*, Vol 73 (1913-14), Part 2

[9] P R Halmos, "The theory of unbiased estimation," *Annals of Math Stat*, Vol 17 (1946), pp 34–43.

[10] W Höffding, "On the distribution of the rank correlation coefficient τ, when the variates are not independent," *Biometrika*, Vol 34 (1947), pp. 183–196

[11] H Hotelling and M. R Pabst, "Rank correlation and tests of significance involving no assumptions of normality," *Annals of Math Stat.*, Vol 7 (1936), pp 29–43.

[12] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, Vol 30 (1938), pp. 81–93

[13] M G Kendall, "Partial rank correlation," *Biometrika*, Vol 32 (1942), pp. 277–283

[14] M. G Kendall, S F H Kendall, and B Babington Smith, "The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times," *Biometrika*, Vol 30 (1939), pp 251–273

[15] J W. Lindeberg, "Über die Korrelation," VI *Skand Matematikerkongres i København*, 1925, pp 437–446

[16] J W Lindeberg, "Some remarks on the mean error of the percentage of correlation," *Nordic Statistical Journal*, Vol. 1 (1929), pp 137–141

[17] H. B Mann, "Nonparametric tests against trend," *Econometrica*, Vol 13 (1945), pp 245–259.

[18] R v Mises, "On the asymptotic distribution of differentiable statistical functions," *Annals of Math. Stat.*, Vol 18 (1947), pp 309–348

[19] U. S Nair, "The standard error of Gini's mean difference," *Biometrika*, Vol 28 (1936), 428–436

[20] K. Pearson, "On further methods of determining correlation," *Drapers' Company Research Memoirs*, Biometric Series, IV, London, 1907.

[21] V. Volterra, *Theory of Functionals*, Blackie, (authorized translation by Miss M Long), London and Glasgow, 1931

[22] G U Yule and M G. Kendall, *An Introduction to the Theory of Statistics*, Griffin, 11th Edition, London, 1937

OPTIMUM CHARACTER OF THE SEQUENTIAL PROBABILITY RATIO TEST

A. WALD AND J. WOLFOWITZ

Columbia University

1. Summary. Let S_0 be any sequential probability ratio test for deciding between two simple alternatives H_0 and H_1, and S_1 another test for the same purpose. We define $(i, j = 0, 1)$:

$\alpha_i(S_j)$ = probability, under S_j, of rejecting H_i when it is true;

$E_i^j(n)$ = expected number of observations to reach a decision under test S_j when the hypothesis H_i is true. (It is assumed that $E_i^1(n)$ exists.)

In this paper it is proved that, if

$$\alpha_i(S_1) \leq \alpha_i(S_0) \qquad (i = 0,1),$$

it follows that

$$E_i^0(n) \leq E_i^1(n) \qquad (i = 0, 1).$$

This means that of all tests with the same power the sequential probability ratio test requires on the average fewest observations. This result had been conjectured earlier ([1], [2])

2. Introduction. Let $p_i(x)$, $i = 0, 1$, denote two different probability density functions or (discrete) probability functions. (Throughout this paper the index i will always take the values 0, 1). Let X be a chance variable whose distribution can only be either $p_0(x)$ or $p_1(x)$, but is otherwise unknown. It is required to decide between the hypotheses H_0, H_1, where H_i states that $p_i(x)$ is the distribution of X, on the basis of n independent observations $x_1, \cdots, x_n$ on X, where n is a chance variable defined (finite) on almost every infinite sequence

$$\omega = x_1, x_2, \cdots$$

i e., n is finite with probability one according to both $p_0(x)$ and $p_1(x)$. The definition of $n(\omega)$ together with the rule for deciding on H_0 or H_1 constitute a sequential test

A sequential probability ratio test is defined with the aid of two positive numbers, $A^* > 1$, $B^* < 1$, as follows: Write for brevity

$$p_{ij} = \prod_{k=1}^{j} p_i(x_k).$$

Then $n = j$ if

$$\frac{p_{1j}}{p_{0j}} \geq A^* \qquad \text{or} \qquad \leq B^*$$

and

$$B^* < \frac{p_{1k}}{p_{0k}} < A^*, \qquad k < j.$$

If

$$\frac{p_{1n}}{p_{0n}} \geq A^*, \qquad \text{the hypothesis } H_1 \text{ is accepted,}$$

if

$$\frac{p_{1n}}{p_{0n}} \leq B^* \text{ the hypothesis } H_0 \text{ is accepted}$$

In this paper we limit consideration to sequential tests for which $E_i(n)$ exists, where $E_i(n)$ is the expected value of n when H_i is true (i e., when $p_i(x)$ is the distribution of X). It has been proved in [3] that all sequential probability ratio tests belong to this class. The purpose of the paper is to prove the result stated in the first section. Throughout the proof we shall find it convenient to assume that there is an a priori probability g_i that H_i is true ($g_0 + g_1 = 1$, we shall write $g = (g_0, g_1)$). We are aware of the fact that many statisticians believe that in most problems of practical importance either no a priori probability distribution exists, or that even where it exists the statistical decision must be made in ignorance of it; in fact we share this view. Our introduction of the a priori probability distribution is a purely technical device for achieving the proof which has no bearing on statistical methodology, and the reader will verify that this is so. We shall always assume below that $g_0 \neq 0, 1$

Let W_0, W_1, c be given positive numbers. We define

$$R = g_0(W_0\alpha_0 + cE_0(n)) + g_1(W_1\alpha_1 + cE_1(n)),$$

and call R the average risk associated with a test S and a given g (obviously R is a function of both). We shall say that H_i is accepted when the decision is made that $p_i(x)$ is the distribution of X. We shall say that H_0 is rejected when H_1 is accepted, and vice versa. The reader may find it helpful to regard W_i as a weight which measures the loss caused by rejecting H_i when it is true, c as the cost of a single observation, and R as the average loss associated with a given g and a test S. For mathematical purposes these are simply quantities which we manipulate in the course of the proof.

3. Role of the probability ratio. Let g, $W = (W_0, W_1)$, and c be fixed. Let S be a given sequential test, with $R(S)$ the associated risk and $n(\omega, S)$ the associated "sample size" function. Let $\psi(x_1, \cdots, x_n)$ be the "decision" function; this is a function which takes only the values 0 and 1, and such that, when $x_1, \cdots, x_n$ is the sample point, the hypothesis with index $\psi(x_1, \cdots, x_n)$ is *rejected*. Define the following decision function $\varphi(x_1, \cdots, x_n)$: $\varphi = 0$ when

$$\lambda = \frac{W_1 g_1 p_{1n}}{W_0 g_0 p_{0n}}$$

is greater than 1, and $\varphi = 1$ when $\lambda < 1$ When $\lambda = 1$, φ may be 0 or 1 at pleasure

It must be remembered that an actual decision function is a single-valued function of $(x_1, \cdots, x_n)$. We note, however, that

a) the relevant properties of a test are not affected by changing the test on a set T' of points ω whose probability is zero according to both H_0 and H_1, i.e., changing the definition on T of n and/or of the decision function, leaves α_0, α_1, $E_0(n)$ and $E_1(n)$ unaltered In particular, the average risk R remains unchanged.

b) the set of points for which $p_{0n} = p_{1n} = 0$ and λ is indeterminate, has probability zero according to both H_0 and H_1.

In view of the above we decide arbitrarily, in all sequential tests which we shall henceforth consider, to define $n = j$, and $\psi = 0$, whenever $p_{0j} = p_{1j} = 0$, and $n \neq 1, \cdots, (j - 1)$. By this arbitrary action $R(S)$ will not be changed.

Let now

$$L_{in} = \frac{W_i g_i p_{in}}{g_0 p_{0n} + g_1 p_{1n}};$$

$$L_n = cn + \min (L_{0n}, L_{1n}).$$

We have

$$EL_{\psi n} = \Sigma g_i W_i \alpha_i$$

where the operator E denotes the expected value with respect to the joint distribution of H_i and $(x_1, \cdots, x_n)$, i.e., E is the operator $g_0E_0 + g_1E_1$. If now the event $\{\psi(S) \neq \varphi \text{ and } \lambda \neq 1\}$ has positive probability according to either H_0 or H_1, we would have, for $n = n(\omega, S)$,

$$EL_{\varphi n} < EL_{\psi n}$$

Hence, if the decision function ψ connected with the test S were replaced by the decision function φ, R would be decreased Since our object throughout this proof will be to make R as small as possible, we shall confine ourselves henceforth, except when the contrary is explicitly stated, to tests for which φ is the decision function. This will be assumed even if not explicitly stated

The function φ has not yet been uniquely defined when $\lambda = 1$ A definition convenient for later purposes will be given in the next section. R is the same for all definitions.

We thus have that φ is a function only of λ, or, what comes to the same thing when W is fixed, of $r_n = \dfrac{p_{1n}}{p_{0n}}$ Define

$$r_j = \frac{p_{1j}}{p_{0j}}, \qquad j = 1, 2, \cdots$$

We shall now prove

LEMMA 1. *Let g, W, and c be fixed. There exists a sequential test S^* for which the average risk is a minimum. Its sample size function $n(\omega, S^*)$ can be defined by means of a properly chosen subset K of the non-negative half-line as follows. For any ω consider the associated sequence*

$$r_1, r_2, \cdots$$

and let j be the smallest integer for which $r_j \in K$. Then $n = j$. The function n may be undefined on a set of points ω whose probability according to H_0 and H_1 is zero.

Let $a = (a_1, \cdots, a_d)$ be any point in some finite d-dimensional Euclidean space, provided only that $p_{0d}(a)$ and $p_{1d}(a)$ are not both zero. Let $b = \dfrac{p_{1d}(a)}{p_{0d}(a)}$ and let $l(a) = cd + \min(L_{0d}, L_{1d})$. Let D be any sequential test whatever for which $n(\omega, D) > d$ for any ω whose first d coordinates are the same as those of a, and for which $E(n \mid a, D) < \infty$, where $E(n \mid a, D)$ is the conditional expected value of n according to the test D under the condition that the first d coordinates of ω are the same as those of a. For brevity let G represent the set of points ω which fulfill this last condition, i.e., that the first d coordinates of ω are the same as those of a. Finally, let $E(L_n \mid a, D)$ be the conditional expected value of L_n according to D under the condition that ω is in the set G. We know that $\min(L_{0d}, L_{1d})$ depends only on $r_d(a) = b$.

Write

$$\nu(a) = \sup_D [l(a) - E(L_n \mid a, D)].$$

Let $a_0 = (a_{01}, \cdots, a_{0k})$ be any point such that

$$\frac{p_{1d}(a)}{p_{0d}(a)} = \frac{p_{1k}(a_0)}{p_{0k}(a_0)}.$$

Let D_0 be any sequential test whatever for which $n(\omega, D_0) > k$ for any ω whose first k coordinates are the same as those of a_0, and for which $E(n \mid a_0, D_0) < \infty$. Let

$$\nu(a_0) = \sup_{D_0} [l(a_0) - E(L_n \mid a_0, D_0)].$$

We shall prove that $\nu(a) = \nu(a_0)$. Thus we shall be justified in writing

$$\gamma(b) = \nu(a) = \nu(a_0)$$

Suppose, therefore that $\nu(a) > \nu(a_0)$. Let D_1 be a test of the type D such that

$$l(a) - E(L_n \mid a, D_1) > \frac{\nu(a) + \nu(a_0)}{2}.$$

We now partially define another sequential test D_{10} of the type D_0 as follows: Let

$$\bar{a} = a_1, \cdots, a_d, y_1, \cdots, y_t,$$

be any sequence such that $n(\bar{a}, D_1) = d + t$. Then for the sequence

$$\bar{a}_0 = a_{01}, \cdots, a_{0k}, y_1, \cdots, y_t,$$

let $n(\bar{a}_0, D_{10}) = k + t$. The decision function ψ_0 associated with D_{10} will be partially defined as follows:

$$\psi_0(\bar{a}_0) = \varphi(\bar{a}).$$

(The reader will observe that it may happen that $\psi_0(\bar{a}_0) \neq \varphi(\bar{a}_0)$). Since $r_d(a) = r_k(a_0)$ it follows that

$$l(a) - E(L_n \mid a, D_1) = l(a_0) - E(L_n \mid a_0, D_{10}) > \frac{\nu(a) + \nu(a_0)}{2} > \nu(a_0),$$

in violation of the definition of $\nu(a_0)$. A similar contradiction is obtained if $\nu(a) < \nu(a_0)$ Hence $\nu(a) = \nu(a_0)$ as was stated above.

We define K to consist of all numbers b which are such that there exist points a with $r_d(a) = b$, and for which $\gamma(b) \leq 0$. We shall now prove that the test S^* defined in the statement of the lemma is such that $R(S^*)$ is a minimum. Recall that the average risk is the expected value of L_n Let S be any other test. Let $a^* = (\overset{*}{a_1}, \cdot \quad, \overset{*}{a_{d^*}})$ be any sequence such that either $n(a^*, S^*) = d^*$, or $n(a^*, S) = d^*$, but $n(a^*, S^*) \neq n(a^*, S)$. We exclude the trivial case that the probability of the occurrence of such a sequence, under both H_0 and H_1, is zero. Let $r_{d^*}(a^*) = b^*$. The sequence a^* may be one of three types:

1) $\gamma(b^*) < 0$. Hence $b^* \in K$, $n(a^*, S) > d^*$. It is more advantageous, from the point of view of diminishing the average risk, to terminate the sequential process at once, since $E(L_n \mid a^*, S) > l(a^*)$.

2) $\gamma(b^*) = 0$. Hence $b^* \in K$, $n(a^*, S) > d^*$. If $l(a^*) - E(L_n \mid a^*, S) = 0$, i.e., the supremum is actually attained by S, then, as far as the average risk is concerned, it makes no difference whether the sequential process is terminated with a^* or continued according to S. If, however, $l(a^*) - E(L_n \mid a^*, S) < 0$, it is clearly disadvantageous to proceed according to S. It is impossible that $l(a^*) - E(L_n \mid a^*, S) > 0$, since $\gamma(b^*) = 0$.

3) $\gamma(b^*) > 0$. Hence $b^* \notin K$, $n(a^*, S) = d^*$. Clearly it is more advantageous from the point of view of diminishing the average risk not to terminate the sequential process, but to continue with at least one more observation. After one more observation we are either in case 1 or 2, where it is advantageous to terminate the sequential process, or again in case 3, where it is advantageous to take yet another observation.

We conclude that $R(S^*)$ is a minimum, as was to be proved

4. A fundamental lemma. Consider the complement of K with respect to the non-negative half-line, and from it delete all points b' for which there exists no point a in some d-dimensional Euclidean space such that $r_d(a) = b'$. The point 1 is never to be considered as of the type of b', i.e., 1 is never to be deleted Designate the resulting set by $\bar{K}$.

Our proof of the theorem to which this paper is devoted hinges on the following lemma:

LEMMA 2. *Let W, g, c be fixed, and $\bar{K}$ be as defined above. There exist two positive numbers A and B, with $B \leq \frac{W_0 g_0}{W_1 g_1} \leq A$, such that*

a) if $b \epsilon K$, then either $b \geq A$ or $b \leq B$

b) if $b \epsilon \bar{K}$, $B \leq b \leq A$.

Two remarks may be made before proceeding with the proof:

1) We may now complete the definition of φ for tests of the type of S^*. The reader will recall that φ was not uniquely defined when $\lambda = 1$, i.e., when $r_n = \frac{W_0 g_0}{W_1 g_1}$. Lemma 2 shows that it is necessary to define $\varphi(\lambda)$ only when $\lambda = \frac{W_0 g_0}{W_1 g_1} \epsilon K$ and λ is therefore either A or B. We will define $\varphi\left(\frac{W_0 g_0}{W_1 g_1}\right)$ as 0 or 1, according as $\frac{W_0 g_0}{W_1 g_1}$ is A or B, and $A \neq B$. This is simply a convenient definition which will give uniqueness. When $A = B = \frac{W_0 g_0}{W_1 g_1} \epsilon K$, the situation is completely trivial, and we may take $\varphi = 0$ arbitrarily

2) If $1 \epsilon K$ the above lemma shows that the average risk is minimized (for fixed W, g, c, of course) by taking no observations at all. We have $\varphi = 0$ or 1 according as $1 \geq A$ or $1 \leq B$

PROOF OF THE LEMMA: Let $h > \frac{W_0 g_0}{W_1 g_1}$ be a point in $\bar{K}$. We will prove that any point h' such that $\frac{W_0 g_0}{W_1 g_1} \leq h' < h$, and such that there exists a point a' in some d'-dimensional Euclidean space for which $r_{d'}(a') = h'$, is also in $\bar{K}$ In a similar way it can be shown that, if $h_0 < \frac{W_0 g_0}{W_1 g_1}$ is any point in $\bar{K}$, any point h_0' such that $h_0 < h_0' \leq \frac{W_0 g_0}{W_1 g_1}$, and such that there exists a point a_0' in some d''-dimensional Euclidean space for which $r_{d''}(a_0') = h_0'$, is also in $\bar{K}$ This will prove the lemma

Let therefore h and h' be as above. Let S^* be the sequential test based on K, with the decision function φ. Let a be a point in d-space such that $r_d(a) = h$. Since $h \epsilon \bar{K}$ we have $\gamma(h) > 0$

We now wish to define partially another sequential test $\bar{S}$, with a decision function which may be different from φ, as follows: Let a' be defined as above. Write

$$a = (a_1, \cdots, a_d)$$
$$a' = (a_1', \cdots, a_{d'}').$$

Let

$$\bar{a} = a_1, \cdots, a_d, y_1, \cdots, y_t$$

be any sequence such that $n(\bar{a}, S^*) = d + l$. Then for the sequence

$$\bar{a}' = a_1', \cdots, a_{d'}', y_1, \cdots, y_l$$

let $n(\bar{a}', \bar{S}) = d' + l$ The decision function ψ associated with $\bar{S}$ will be partially defined as follows:

$$\psi(\bar{a}') = \varphi(\bar{a}).$$

Clearly

$$E_i(n \mid a, S^*) - d = E_i(n \mid a', \bar{S}) - d' \qquad (i = 0, 1) \tag{4.1}$$

and

$$E_i(\varphi \mid a, S^*) = E_i(\psi \mid a', \bar{S}) \qquad (i = 0, 1). \tag{4.2}$$

Furthermore, we have

$$\begin{aligned} &l(a) - E(L_n \mid a, S^*) \\ &= \frac{g_0}{g_0 + g_1 h} \{W_0 + cd - cE_0(n \mid a, S^*) - W_0[1 - E_0(\varphi \mid a, S^*)]\} \\ &\quad + \frac{g_1 h}{g_0 + g_1 h} \{cd - cE_1(n \mid a, S^*) - W_1 E_1(\varphi \mid a, S^*)\}. \end{aligned} \tag{4.3}$$

Since $\gamma(h) > 0$, and since

$$cd - cE_1(n \mid a, S^*) - W_1 E_1(\varphi \mid a, S^*) < 0, \tag{4.4}$$

we must have

$$W_0 + cd - cE_0(n \mid a, S^*) - W_0[1 - E_0(\varphi \mid a, S^*)] > 0. \tag{4.5}$$

From $h' < h$ it follows that

$$\frac{g_0}{g_0 + g_1 h'} > \frac{g_0}{g_0 + g_1 h}, \qquad \text{and} \qquad \frac{g_1 h'}{g_0 + g_1 h'} < \frac{g_1 h}{g_0 + g_1 h}. \tag{4.6}$$

Relations (4.1), (4.2), (4.4), (4.5) and (4.6) imply that the value of the right hand member of (4.3) is increased by replacing φ, h, a, S^* and d by ψ, h', a', $\bar{S}$, and d', respectively This proves our lemma.

If there are values which r_j cannot assume the pair B, A might not be unique For convenience we shall define A and B uniquely in the manner described below. We will always adhere to this definition thereafter

We shall first define $\gamma(h)$ for all positive h in a manner consistent with the previous definition, which defined $\gamma(h)$ only for those values of h which could be assumed by r_j Let h be any positive number and $D(h)$ be any sequential test with the following properties

(4.7) there exists a set $Q(h)$ of positive numbers such that $n = j$ if and only if the j-th member of the sequence

$$hr_1, hr_2, hr_3, \cdots$$

is the first element of the sequence to be in $Q(h)$

$$E_i(n \mid D(h)) < \infty \qquad (i = 0, 1). \tag{4.8}$$

We define, for $h \geq \dfrac{W_0 g_0}{W_1 g_1}$,

$$\gamma(h \mid D(h)) = \frac{g_0}{g_0 + g_1 h} \{W_0 E_0(\varphi \mid D(h)) - cE_0(n \mid D(h))\} \tag{4.9}$$
$$+ \frac{g_1 h}{g_0 + g_1 h} \{-W_1 E_1(\varphi \mid D(h)) - cE_1(n \mid D(h))\},$$

$$\gamma(h) = \sup_{D(h)} \gamma(h \mid D(h)) \tag{4.10}$$

with a corresponding definition for $h \leq \dfrac{W_0 g_0}{W_1 g_1}$. Thus $\gamma(h)$ is defined for all positive h. This definition coincides with the previous definition whenever the latter is applicable It is true that the supremum operation in (4 10) is limited to tests which depend only on the probability ratio, as (4 7) implies, but the argument of Lemma 1 shows that this limitation does not diminish the supremum (It might appear that, for $h = \dfrac{W_0 g_0}{W_1 g_1}$, $\gamma(h)$ is not uniquely defined. We shall shortly see that this is not the case.)

The quantity $\gamma(h)$ depends, of course, on g_0 and g_1 To put this in evidence, we shall also write $\gamma(h, g_0, g_1)$. One can easily verify that

$$\gamma(h, g_0, g_1) = \gamma\left(1, \frac{g_0}{g_0 + g_1 h}, \frac{g_1 h}{g_0 + g_1 h}\right).$$

More generally, for any positive values h and h', we have $\gamma(h, g_0, g_1) = \gamma(h', \bar{g}_0, \bar{g}_1)$, where $\bar{g}_0$ and $\bar{g}_1$ are suitable functions of g_0, g_1, h, and h'. Thus, if h is not an admissible value of the probability ratio and h' is any admissible value, we can interpret the value of $\gamma(h, g_0, g_1)$ as the value of γ corresponding to h' and some properly chosen a priori probabilities $\bar{g}_0$ and $\bar{g}_1$

We now define A as the greatest lower bound of all points $h \geq \dfrac{W_0 g_0}{W_1 g_1}$ for which $\gamma(h) \leq 0$ We define B as the least upper bound of all points $h \leq \dfrac{W_0 g_0}{W_1 g_1}$ for which $\gamma(h) \leq 0$ If $\gamma(h) \leq 0$ for all h the above definition implies $A = B = \dfrac{W_0 g_0}{W_1 g_1}$.

The argument of Lemma 2 shows that $\gamma(h)$ is monotonically increasing in the interval $\left(B, \dfrac{W_0 g_0}{W_1 g_1}\right)$, and that $\gamma(h)$ is monotonically decreasing in the interval $\left(\dfrac{W_0 g_0}{W_1 g_1}, A\right)$

We shall now define a sequential test $S^*(h)$ for every positive h The decision

function of $S^*(h)$ will be φ, and $n = j$ if and only if the j-th member of the sequence

$$\gamma(hr_1), \quad \gamma(hr_2), \quad \gamma(hr_3), \cdots$$

is the first element to be ≤ 0. We see that

$$\gamma(h) = \gamma(h \mid S^*(h)) \tag{4.11}$$

for *all* h. Incidentally, this proves that $\gamma(h)$ was uniquely defined at $h = \dfrac{W_0 g_0}{W_1 g_1}$.

We shall now prove

LEMMA 3. *The function $\gamma(h)$ has the following properties*

a) *It is continuous for all h.*

b) $\gamma(A) = \gamma(B) = 0$

c) $\gamma(h) < 0$ *for* $h > A$ *or* $< B$

Only a) and c) require proof, since b) is a trivial consequence of a) and the definition of A and B.

Let h be any point except $\dfrac{W_0 g_0}{W_1 g_1}$, and let z be any point in a neighborhood of h. Within a neighborhood of h both $E_0(n \mid S^*(z))$ and $E_1(n \mid S^*(z))$ are bounded. Let Δ be an arbitrarily given, positive number. Let h' and h'' be any two points in a sufficiently small neighborhood of h, to be described shortly. We proceed as in the argument of Lemma 2, with the present h' corresponding to h of Lemma 2, the present h'' corresponding to h' of Lemma 2, and with $S^*(h')$ corresponding to S^* of Lemma 2. Since $\dfrac{g_0}{g_0 + g_1 z}$ and $\dfrac{g_1 z}{g_0 + g_1 z}$ are continuous functions of z, and since $E_0(n \mid S^*(z))$ and $E_1(n \mid S^*(z))$ are bounded functions of z, we conclude that, when the neighborhood of h is sufficiently small,

$$\gamma(h'') \geq \gamma(h') - \Delta.$$

Reversing the roles of h' and h'' we obtain that in this neighborhood

$$\gamma(h') \geq \gamma(h'') - \Delta,$$

and conclude that

$$|\gamma(h') - \gamma(h'')| \leq \Delta.$$

Since Δ was arbitrary, this implies the continuity of $\gamma(h)$ everywhere, except perhaps at $h = \dfrac{W_0 g_0}{W_1 g_1}$.

To deal with the point $h = \dfrac{W_0 g_0}{W_1 g_1}$, proceed as follows: Using the above argument and the definition (4 9), (4.10), we prove that $\gamma(h)$ is continuous on the right

at $h = \frac{W_0 g_0}{W_1 g_1}$. Using, at the point $h = \frac{W_0 g_0}{W_1 g_1}$, the definition of $\gamma(h \mid D(h))$ for $h \leq \frac{W_0 g_0}{W_1 g_1}$ i.e.,

$$
\begin{aligned}
\gamma(h \mid D(h)) = {} & \frac{g_0}{g_0 + g_1 h} \{-W_0 E_0(1 - \varphi \mid D(h)) - cE_0(n \mid D(h))\} \\
& + \frac{g_1 h}{g_0 + g_1 h} \{W_1 E_1(1 - \varphi \mid D(h)) - cE_1(n \mid D(h))\},
\end{aligned}
\tag{4.12}
$$

(4.10) and (4.11), we prove that $\gamma(h)$ is continuous on the left at $h = \frac{W_0 g_0}{W_1 g_1}$. This proves a).

To prove c), we proceed as follows: Suppose for $h_0 > A$ we had $\gamma(h_0) = 0$ Since

$$\{ -W_1 E_1(\varphi \mid S^*(h_0)) - cE_1(n \mid S^*(h_0))\} < 0,$$

we would have that

$$\{W_0 E_0(\varphi \mid S^*(h_0)) - cE_0(n \mid S^*(h_0))\} > 0.$$

An argument like that of Lemma 2 would then show that $\gamma(h) > 0$ for $\frac{W_0 g_0}{W_1 g_1} < h < h_0$. This, however, is impossible, because it is a violation of the definition of A

In a similar way we prove that if $h < B$, $\gamma(h) < 0$ This proves c) and with it the lemma.

5. The behavior of A and B.

LEMMA 4. *Let g and c be fixed Then A and B are continuous functions of W_0 and W_1*

PROOF: It will be sufficient to prove that A is continuous, the proof for B being similar Suppose $A > B$. Let h_1 and h_2 be such that

a) $B < h_1 < A < h_2$;

b) $h_2 - h_1 < \Delta$ for an arbitrary positive Δ.

We write $\gamma(h)$ temporarily as $\gamma(h, W_0, W_1)$ in order to exhibit the dependence on W_0 and W_1. Then

$$\gamma(h_1, W_0, W_1) > 0,$$
$$\gamma(h_2, W_0, W_1) < 0.$$

It follows from (4.9) that $\gamma(h \mid D(h))$ is continuous in W_0, W_1, uniformly in $D(h)$. Hence $\gamma(h, W_0, W_1) = \sup_{D(h)} \gamma(h \mid D(h))$ is also continuous in W_0, W_1

Hence, for ΔW_0 and ΔW_1 sufficiently small,

$$\gamma(h_1, W_0 + \Delta W_0, W_1 + \Delta W_1) > 0,$$
$$\gamma(h_2, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0.$$

Therefore

$$h_1 \leq A(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq h_2,$$

which proves continuity, since Δ was arbitrary.

If $\frac{W_0 g_0}{W_1 g_1} = A = B$, we take $h_1 < \frac{W_0 g_0}{W_1 g_1} < h_2$, $h_2 - h_1 < \Delta$, and by a similar argument show that

$$\gamma(h_1, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0;$$
$$\gamma(h_2, W_0 + \Delta W_0, W_1 + \Delta W_1) < 0.$$

Thus

$$h_1 \leq B(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq A(W_0 + \Delta W_0, W_1 + \Delta W_1) \leq h_2.$$

This proves the lemma

LEMMA 5 *Let g, c, and W_1 be fixed. A is strictly monotonic in W_0. As W_0 approaches 0, A approaches 0, as W_0 approaches $+\infty$, A also approaches $+\infty$*

PROOF: Since $A \geq \frac{W_0 g_0}{W_1 g_1}$, $A \to +\infty$ as $W_0 \to +\infty$. If $W_0 < c$ no reduction in average risk could compensate for taking even a single observation, no matter what the value of h. Hence $\gamma(h) \leq 0$ for all h when $W_0 < c$, so that $A = B$. Since $B \leq \frac{W_0 g_0}{W_1 g_1}$, $B \to 0$ as $W_0 \to 0$. Hence $A \to 0$ as $W_0 \to 0$. It is evident from (4 9) that $\gamma(h \mid D(h))$ is non-decreasing with increasing W_0 (everything else fixed) Hence also

$$\gamma(h) = \sup_{D(h)} \gamma(h \mid D(h)),$$

is non-decreasing with increasing W_0, for fixed $h > \frac{W_0 g_0}{W_1 g_1}$ and fixed W_1. For a positive Δ sufficiently small and for any h such that $A \leq h < A + \Delta$, we have that

$$E_0(\varphi \mid S^*(h)) > 0.$$

Hence, for such h, $\gamma(h, W_0, W_1)$ is strictly monotonically increasing with increasing W_0 Therefore A is (strictly) monotonically increasing with increasing W_0.

We now define the function $W_0(W_1, \delta)$ of the two positive arguments W_1, δ so that

$$A(W_0(W_1, \delta), W_1) = \delta.$$

By Lemma 5 such a function exists and is single-valued.

6. Properties of the function $W_0(W_1, \delta)$ LEMMA 6. *$W_0(W_1, \delta)$ is continuous in W_1*

PROOF: Let

$$\lim_{N \to \infty} W_{1N} = W_1,$$

and suppose that the sequence $\{W_0(W_{1N}, \delta)\}$ did not converge. Suppose W_0' and W_0'' were two distinct limit points of this sequence. From the continuity of A (Lemma 4) it would follow that

$$A(W_0', W_1) = A(W_0'', W_1)$$

This, however, violates Lemma 5. The only remaining possibility to be considered is that

$$\lim_{N \to \infty} W_0(W_{1N}, \delta) = \infty$$

If that were the case, then, since $A \geq \frac{W_0 g_0}{W_1 g_1}$, it would follow that $A \to \infty$, in violation of the fact that $A \equiv \delta$

LEMMA 7. *We have, for fixed* δ,

$$\lim_{W_1 \to 0} W_0(W_1) = 0,$$

$$\lim_{W_1 \to \infty} W_0(W_1) = \infty$$

PROOF: If, for small W_1, $W_0(W_1)$ were bounded below by a positive number, then, since $A \geq \frac{g_0 W_0(W_1, \delta)}{W_1 g_1}$, we could make A arbitrarily large by taking W_1 sufficiently small, in violation of the fact that $A \equiv \delta$ To prove the second half of the lemma, assume that $W_0(W_1)$ is bounded above as $W_1 \to \infty$. Then $B\left(\leq \frac{W_0 g_0}{W_1 g_1}\right)$ will approach zero as $W_1 \to \infty$ Let h be fixed so that $B < h < \delta$. Consider the totality of points ω for which there exists an integer $n^*(\omega)$ such that:

$$hr_{n^*} \leq B,$$

$$B < hr_j < \delta, \qquad j < n^*.$$

The conditional expected value of n^* in this totality, when H_0 is true, may be made arbitrarily large by making B sufficiently small. Hence, when W_1 is sufficiently large, for fixed but arbitrary $h < \delta$, the optimum procedure from the point of minimizing the average risk is to reject H_0 at once without taking any more observations This, however, contradicts the fact that $h < \delta$, and proves the lemma

LEMMA 8. *We have, for fixed* $\delta > 1$,

$$\lim_{W_1 \to 0} B(W_0(W_1, \delta), W_1) = \delta;$$

$$\lim_{W_1 \to \infty} B(W_0(W_1, \delta), W_1) = 0.$$

PROOF: By Lemma 7,

$$\lim_{W_1 \to 0} W_0(W_1) = 0$$

When, for fixed c, both W_0 and W_1 are small enough, then, no matter what the value of h, $\gamma(h) < 0$. Hence $A = B$, which proves the first half of the lemma.

Let now $\{W_{1N}\}$ be a sequence such that $\lim W_{1N} = \infty$ Let $\delta > 1$ For the sake of brevity we write $B(W_{1N})$ instead of

$$B(W_0(W_{1N}\delta), W_{1N}).$$

Suppose that, for sufficiently large N, $B(W_{1N})$ is bounded below by a positive number. Hence, for sufficiently large N, the probability of rejecting H_1 when it it is true is bounded below by a positive number. Moreover, since $B \leq \frac{W_0 g_0}{W_1 g_1} \leq A$, it follows that, for N sufficiently large, $\frac{W_{0N} g_0}{W_{1N} g_1}$ is bounded above and below by positive constants Thus, for large N the average risk of the test defined by $B(W_{1N})$, δ, is greater than ug_1W_{1N}, where u is a positive constant which does not depend on N Moreover, from the definition of $B(W_{1N})$, this risk is a minimum

Let ϵ be a positive number such that $\epsilon\left(\frac{W_{0N} g_0}{W_{1N} g_1} + 1\right) < \frac{u}{2}$ for all N sufficiently large. Let V_1, V_2, with $0 < V_1 < 1 < V_2$, be two constants such that, for the sequential probability ratio test determined by them, both α_0 and α_1 are $< \epsilon$. Of course $E_0 n$ and $E_1 n$ are finite and determined by the test. For this test the average risk is less than

$$\begin{aligned} &\epsilon(g_0 W_{0N} + g_1 W_{1N}) + cg_0 E_0 n + cg_1 E_1 n \\ &< \frac{u}{2} g_1 W_{1N} + cg_0 E_0 n + cg_1 E_1 n \\ &< \frac{3u}{4} g_1 W_{1N}, \end{aligned}$$

for W_{1N} large enough. This however contradicts the fact that the minimum risk is $> ug_1W_{1N}$, and proves the lemma.

7. Proof of the theorem. Let a given sequential probability ratio test S_0 be defined by B^*, A^*; $B^* < 1 < A^*$. Let $\alpha_i(S_0)$ be the probability, according to S_0, of rejecting H_i when it is true. Let c be fixed.

By Lemma 4, B is a continuous function of W_0 and W_1. Let $\delta = A^*$ in Lemma 8. Then there exists a pair $\overline{W}_0$, $\overline{W}_1$, with $\overline{W}_0 = W_0(\overline{W}_1, A^*)$, such that

$$A(\overline{W}_0, \overline{W}_1) = A^*,$$
$$B(\overline{W}_0, \overline{W}_1) = B^*.$$

Hence the average risk

$$\sum_i g_i [\overline{W}_i \alpha_i(S_0) + cE_i^0(n)],$$

corresponding to the sequential test S_0 is a minimum.

Now let S_1 be any other test for deciding between H_0 and H_1 and such that

$$\alpha_i(S_1) \le \alpha_i(S_0), \quad \text{and} \quad E_i^1(n) \text{ exists } (i = 1, 2).$$

Then

$$\sum_i g_i [\overline{W}_i \alpha_i(S_0) + cE_i^0(n)] \le \sum_i g_i [\overline{W}_i \alpha_i(S_1) + cE_i^1(n)].$$

Since $\alpha_i(S_1) \le \alpha_i(S_0)$, we have

$$\sum_i g_i E_i^0(n) \le \sum_i g_i E_i^1(n).$$

Now g_0, g_1 were arbitrarily chosen (subject, of course, to the obvious restrictions). Hence it must be that

$$E_i^0(n) \le E_i^1(n).$$

This, however, is the desired result.

REFERENCES

[1] A. Wald, "Sequential tests of statistical hypotheses", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 117–186.
[2] A. Wald, *Sequential Analysis*, John Wiley and Sons, Inc., New York, 1947.
[3] Charles Stein, "A note on cumulative sums", *Annals of Math. Stat.*, Vol. 17 (1946), pp. 498–499.

LIMITING DISTRIBUTION OF A ROOT OF A DETERMINANTAL EQUATION

By D. N. Nanda

Institute of Statistics, University of North Carolina

1. Summary. The exact distribution of a root of a determinantal equation when the roots are arranged in a monotonic order was obtained by S. N. Roy [3] in 1943. A different method for deriving the distribution of any one of these roots has been described by the author in [2]. In the present paper the limiting forms of these distributions are obtained. This paper gives a method by which the limiting distributions can be obtained without undergoing an inordinate amount of mathematical labor.

2. Introduction. If $x = \| x_{ij} \|$ and $x^* = \| x^*_{ij} \|$ are two p-variate sample matrices with n_1 and n_2 degrees of freedom and $S = \| xx' \|/n_1$ and $S^* = \|x^*x^{*\prime}\|/n_2$ are the covariance matrices which under the null hypothesis are independent estimates of the same population covariance matrix, then the joint distribution of the roots of the determinantal equation $| A - \theta(A + B) | = 0$, where $A = n_1S$ and $B = n_2S^*$, was obtained by Hsu [1] in 1939 and is

$$(1) \quad R'(l, \mu, \nu) = \frac{\pi^{l/2} \prod_{i=1}^{l} \Gamma\left(\frac{l+\mu+\nu+i-2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{\mu+i-1}{2}\right) \Gamma\left(\frac{\nu+i-1}{2}\right) \Gamma\left(\frac{i}{2}\right)} \prod_{i=1}^{l} (\theta_i)^{\mu/2-1} \prod_{i=1}^{l} (1-\theta_i)^{\nu/2-1} \prod_{i<j} (\theta_i - \theta_j),$$

$$(0 \leq \theta_l \leq \theta_{l-1} \leq \cdots \leq \theta_1 \leq 1),$$

where $l = \min(p, n_1)$, $\mu = | p - n_1 | + 1$ and $\nu = n_2 - p + 1$. The distribution density may be expressed as

$$(2) \qquad R(l, m, n) = c(l, m, n) \prod_{i=1}^{l} [\theta_i^m (1-\theta_i)^n] \prod_{i<j} (\theta_i - \theta_j),$$

where $m = \mu/2 - 1$ and $n = \nu/2 - 1$.

3. Method. Let $\theta_i = \zeta_i/n$ in (2). The joint distribution reduces to

$$(3) \qquad \frac{c(l, m, n)}{n^{l+lm+l(l-1)/2}} \prod_{i=1}^{l} [\zeta_i^m (1 - \zeta_i/n)^n] \prod_{i<j} (\zeta_i - \zeta_j)\, d\zeta_1 \cdots d\zeta_l,$$

$$(0 \leq \zeta_l \leq \zeta_{l-1} \cdots \leq \zeta_1 \leq n).$$

As n tends to infinity the limit of (3) is

$$(4) \qquad K(l, m) \quad \prod_{i=1}^{l} \zeta_i^m \prod_{i<j} (\zeta_i - \zeta_j) e^{-\Sigma \zeta_i} \, d\zeta_i$$

$$(0 \leq \zeta_l \leq \zeta_{l-1} \cdots \leq \zeta_1 < \infty).$$

The value of $K(l, m)$ is

$$\lim_{n \to \infty} \frac{c(l, m, n)}{n^{l+lm+l(l-1)/2}}$$

$$= \lim_{n\to\infty} \frac{\pi^{l/2} \prod_{i=1}^{l} \Gamma\left(\frac{l + 2m + 2n + i + 2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{2m + i + 1}{2}\right) \Gamma\left(\frac{2n + i + 1}{2}\right) \Gamma(i/2) \cdot n^{l+lm+l(l-1)/2}}$$

$$= \frac{\pi^{l/2}}{\prod_{i=1}^{l} \Gamma\left(\frac{2m + i + 1}{2}\right) \Gamma(i/2)} \lim_{n\to\infty} \frac{\prod_{i=1}^{l} \Gamma\left(\frac{l + 2m + 2n + i + 2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{2n + i + 1}{2}\right) \cdot n^{l+lm+l(l-1)/2}}$$

By using Stirling's approximation for gamma functions and after simplification we get

$$\lim_{n\to\infty} \frac{\prod_{i=1}^{l} \Gamma\left(\frac{l + 2m + 2n + i + 2}{2}\right)}{\prod_{i=1}^{l} \Gamma\left(\frac{2n + i + 1}{2}\right) \quad n^{lm+l(l+1)/2}} = 1$$

Hence

$$K(l, m) = \frac{\pi^{l/2}}{\prod_{i=1}^{l} \Gamma\left(\frac{2m + i + 1}{2}\right) \Gamma(i/2)},$$

and therefore

$$(5) \qquad \begin{aligned} K(2, m) &= 2^{2m+1}/\Gamma(2m + 2), \\ K(3, m) &= 2^{2m+3}/[\Gamma(m + 1)\Gamma(2m + 3)], \\ K(4, m) &= 2^{4m+5}/[\Gamma(2m + 2)\Gamma(2m + 4)], \\ K(5, m) &= 2^{4m+9}/[3\Gamma(m + 1)\Gamma(2m + 3)\Gamma(2m + 5)]. \end{aligned}$$

Let

$$(6) \qquad G_{l,m}(x) = K(l, m) \int_{0 \leq \zeta_l \leq \zeta_{l-1} \cdots < \zeta_1 \leq x} \prod_{i=1}^{l} \zeta_i^m \prod_{i<j} (\zeta_i - \zeta_j) e^{-\Sigma \zeta_i} \prod d\zeta_i .$$

It can easily be observed that

$$G_{l,m}(x) = \Pr(\zeta_1 \leq x) = \lim_{n\to\infty} \Pr(n\theta_1 \leq x) = \lim_{n\to\infty} \Pr\left(\theta_1 \leq \frac{x}{n}\right).$$

Thus the limiting form of the distribution of the largest root can be obtained by integrating the density given in (4) according to the method described by the author in [2]. It is, however, observed that the mathematical labor is reduced considerably by adopting the following method.

Referring to the results of the exact distribution of the largest root given in [2], let $F_{l,m,n}(x) = (0, l, l-1, \cdots, 1, x, m, n)$, thus $F_{2,m,n}(x) = (0, 2, 1, x, m, n)$ and $F_{3,m,n}(x) = (0, 3, 2, 1, x; m, n)$. Then $c(l, m, n)F_{l,m,n}(x)$ is the probability that none of the roots θ_i exceeds x, and is thus the cumulative distribution function of the greatest root. We shall show that $\lim_{n\to\infty} c(l, m, n)F_{l,m,n}(x/n) = G_{l,m}(x)$. The reader is, however, asked to refer to [2] for the detailed explanation of the notations and certain mathematical operations used in this paper.

4. Limiting distribution of the largest root. We will derive the distribution of the largest root for $l = 2$ and 3 by the two methods. A straightforward method will be named A. A second method, which proves to be very simple and easy will be called Method B.

(a) $l = 2$

(i) METHOD A. We have,

$$\Pr(n\theta_1 \leq x) = G_{2,m}(x) = K(2, m) \int_{0<\zeta_2<\zeta_1<x} (\zeta_1\zeta_2)^m(\zeta_1 - \zeta_2)e^{-(\zeta_1+\zeta_2)}\, d\zeta_1\, d\zeta_2$$

By using the method described in [2], we have

$$G_{2,m}(x) = K(2, m)\left\{\int_{0<\zeta_2<\zeta_1<x} - \int_{0<\zeta_1<\zeta_2<x} \zeta_2^m e^{-\zeta_2}\cdot\zeta_1^{m+1}e^{-\zeta_1}\, d\zeta_1\, d\zeta_2\right\},$$

$$= K(2, m)\{T_0^{m,x}(y, 1, x, m+1) - T_0^{m,x}(0, 1, y; m+1)\},$$

where

$$T_a^{m,b}g(y) = \int_a^b g(y)\cdot y^m e^{-y}\, dy,$$

and

$$(7)(a, 1, b; m+1) = \int_a^b \zeta^{m+1}e^{-\zeta}\, d\zeta = (a^{m+1}e^{-a} - b^{m+1}e^{-b}) + (m+1)(a, 1, b, m).$$

Hence,

$$G_{2,m}(x) = K(2, m)T_0^{m,x}[y^{m+1}e^{-y} - x^{m+1}e^{-x} + (m+1)(y, 1, x; m) + y^{m+1}e^{-y}$$
$$- (m+1)(0, 1, y; m)],$$
$$= K(2, m)T_0^{m,x}[2y^{m+1}e^{-y} - x^{m+1}e^{-x}],$$

as $T_0^{m,x}[(y, 1, x; m) - (0, 1, y; m)] = 0$.

Therefore

$$\lim_{n\to\infty} \Pr\,(n\theta_1 \le x) = G_{2,m}(x) = K(2, m) \tag{8}$$
$$\cdot \left\{2\int_0^x y^{2m+1}e^{-2y}dy - x^{m+1}e^{-x}\int_0^x y^m e^{-y}\,dy\right\}.$$

When $x = \infty$, $G_{2,m}(x) = 1$; hence $K(2, m) = 2^{2m+1}/\Gamma(2m + 2)$, the value given in (5).

Now we shall derive the result by Method B.

(ii) METHOD B.

$$F_{2,m,n}(x) = (0, 2, 1, x; m, n) = \frac{1}{m + n + 2}$$
$$\cdot \left\{2\int_0^x y^{2m+1}(1 - y)^{2n+1}\,dy - x^{m+1}(1 - x)^{n+1}\int_0^x y^m(1 - y)^n\,dy\right\},$$

a result given in [2].

Replacing x by x/n, we get

$$(0, 2, 1, x/n, m, n) = \frac{1}{m + n + 2}$$
$$\cdot \left\{2\int_0^{x/n} y^{2m+1}(1 - y)^{2n+1}\,dy - (x/n)^{m+1}(1 - x/n)^{n+1}\int_0^{x/n} y^m(1 - y)^n\,dy\right\},$$

also, letting $y = u/n$, we have

$$(0, 2, 1, x/n, m, n) = \frac{1}{(m + n + 2)n^{2m+2}} \tag{9}$$
$$\left\{2\int_0^x u^{2m+1}(1 - u/n)^{2n+1}\,du - x^{m+1}(1 - x/n)^{n+1}\int_0^x u^m(1 - u/n)^n\,du\right\}.$$

Thus

$$\lim_{n\to\infty} \Pr\,(n\theta_1 \le x) = \Pr\,(\theta_1 \le x/n) = \lim_{n\to\infty} c(2, m, n)(0, 2, 1, x/n; m, n),$$
$$= \frac{2^{2m+1}}{\Gamma(2m + 2)}\left\{2\int_0^x u^{2m+1}e^{-2u}\,du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\,du\right\},$$

which is the same as (8), obtained by Method A.

(b) $l = 3$.

(i) METHOD A. We have

$$\Pr(n\theta_1 \le x) = G_{3,m}(x) = K(3, m)\int_{0<\zeta_3<\zeta_2<\zeta_1<x} \Pi\zeta_i^m \Pi(\zeta_i - \zeta_j)e^{-\Sigma\zeta_i}\Pi\,d\zeta_i$$
$$= K(3, m)\int_{0<\zeta_3<\zeta_2<\zeta_1<x} (\zeta_1\zeta_2\zeta_3)^m e^{-(\zeta_1+\zeta_2+\zeta_3)}\{1, 2, 3\}\,d\zeta_1\,d\zeta_2\,d\zeta_3,$$

where $\{1, 2, 3\} = \zeta_1\zeta_2\{1, 2\} + \zeta_3\zeta_1\{3, 1\} + \zeta_2\zeta_3\{2, 3\}$, as given in [2].

Or

$$G_{3,m}(x) = K(3, m)\left\{\int_{0<\zeta_3<\zeta_2<\zeta_1<x} + \int_{0<\zeta_1<\zeta_3<\zeta_2<x} + \int_{0<\zeta_2<\zeta_1<\zeta_3<x} \zeta_3^m e^{-\zeta_3}(\zeta_1\zeta_2)^{m+1}e^{-(\zeta_1+\zeta_2)}\{1, 2\}\, d\zeta_1\, d\zeta_2\, d\zeta_3\right\},$$

$$= K(3, m)\{T_0^{m,x}(y, 2, 1, x; m+1) + T_0^{m,x}(0, 2, y, 1, x; m+1) + T_0^{m,x}(0, 2, 1, y; m+1)\},$$

where

$$(a, 2, 1, b, m) = \int_{a<\zeta_2<\zeta_1<b} (\zeta_1\zeta_2)^m(\zeta_1-\zeta_2)e^{-(\zeta_1+\zeta_2)}\, d\zeta_1\, d\zeta_2.$$

We have already obtained

$$(0, 2, 1, x; m) = G_{2,m}(x)/K(2, m) = \left\{2\int_0^x y^{2m+1}e^{-2y}\, dy - x^{m+1}e^{-x}\int_0^x y^m e^{-y}\, dy\right\}$$

as given in (8).

We also need the following results which are obtained by the method described for $l = 2$.

(10) $$(a, 2, 1, b; m) = \left\{2\int_a^b u^{2m+1}e^{-2u}\, du - (a^{m+1}e^{-a} + b^{m+1}e^{-b})\int_a^b u^m e^{-u}\, du\right\},$$

and

(11) $$(a, 2, b, 1, c; m) = \left\{b^{m+1}e^{-b}\int_a^c u^m e^{-u}\, du - a^{m+1}e^{-a}\int_b^c u^m e^{-u}\, du - c^{m+1}e^{-c}\int_a^b u^m e^{-u}\, du\right\}.$$

Using these results we have

$$G_{3,m}(x) = K(3, m)T_0^{m,x}\left\{2\int_y^x u^{2m+3}e^{-2u}\, du - (y^{m+2}e^{-y} + x^{m+2}e^{-x})\int_y^x u^{m+1}e^{-u}\, du - y^{m+2}e^{-y}\int_0^x u^{m+1}e^{-u}\, du + x^{m+2}e^{-x}\int_0^y u^{m+1}e^{-u}\, du + 2\int_0^y u^{2m+3}e^{-2u}\, du - y^{m+2}e^{-y}\int_0^y u^{m+1}e^{-u}\, du\right\}.$$

Simplifying we get

(12) $$\lim_{n\to\infty} \Pr(n\theta_1 \le x) = G_{3,m}(x) = K(3, m)\left\{2\int_0^x u^{2m+3}e^{-2u}\, du\int_0^x u^m e^{-u}\, du - 2\int_0^x u^{m+1}e^{-u}\, du\int_0^x u^{2m+2}e^{-2u}\, du - x^{m+2}e^{-x}\left[2\int_0^x u^{2m+1}e^{-2u}\, du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\, du\right]\right\},$$

where $K(3, m) = 2^{2m+3}/[\Gamma(2m+1)\Gamma(2m+3)]$.

(ii) METHOD B.

$$F_{3,m,n}(x) = (0, 3, 2, 1, x; m, n)$$
$$= \frac{1}{m+n+3}[2(0, 1, x; 2m+3, 2n+1)(0, 1, x; m, n)$$
$$-2(0, 1, x; m+1, n)(0, 1, x; 2m+2, 2n+1)$$
$$-(0, x; m+2, n+1)(0, 2, 1, x; m, n)],$$

a result given in [2].

Replacing x by x/n and putting u/n for the variate y of integration, we have,

$$F_{3,m,n}(x) = (0, 3, 2, 1, x/n; m, n) = \frac{1}{m+n+3}$$
$$\left\{\frac{2}{n^{3m+5}}\int_0^x u^{2m+3}(1-u/n)^{2n+1}\,du\int_0^x u^m(1-u/n)^n\,du - \frac{2}{n^{3m+5}}\right.$$
$$\int_0^x u^{m+1}(1-u/n)^n\,du\int_0^x u^{2m+2}(1-u/n)^{2n+1}\,du - \frac{x^{m+2}(1-x/n)^{n+1}}{n^{3m+4}(m+n+2)}$$
$$\left.\left[2\int_0^x u^{2m+1}(1-u/n)^{2n+1}\,du - x^{m+1}(1-x/n)^{n+1}\int_0^x u^m(1-u/n)^n\,du\right]\right\}.$$

Hence

$$\lim_{n\to\infty}\Pr(n\theta_1 \le x) = \lim_{n\to\infty}\Pr\left(\theta_1 \le \frac{x}{n}\right) = \lim_{n\to\infty} c(3, m, n)F_{3,m,n}(x)$$
$$= K(3, m)\left\{2\int_0^x u^{2m+3}e^{-2u}\,du\int_0^x u^m e^{-u}\,du - 2\int_0^x u^{2m+2}e^{-2u}\,du\int_0^x u^{m+1}e^{+u}\,du\right.$$
$$\left. - x^{m+2}e^{-x}\left[2\int_0^x u^{2m+1}e^{-2u}\,du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\,du\right]\right\},$$

where

$$K(3, m) = 2^{2m+3}/[\Gamma(m+1)\Gamma(2m+3)].$$

This result is the same as (12) obtained by Method A.

We have thus shown that Method B is applicable for obtaining the limiting forms of the distribution of the largest root and that it is much simpler as compared to the straightforward method called Method A here

The limiting distributions for the largest root for $l = 4$ and 5 are listed below.

(c) $l = 4$.

$$\lim_{n\to\infty}\Pr(n\theta_1 \le x) = \lim_{n\to\infty}\Pr\left(\theta_1 \le \frac{x}{n}\right) = G_{4,m}(x)$$
$$= K(4, m)\left\{2\int_0^x u^{2m+5}e^{-2u}\,du\,\frac{G_{2,m}(x)}{K(2, m)} - 2\int_0^x u^{2m+4}e^{-2u}\,du\right.$$
$$\left[2\int_0^x u^{2m+2}e^{-2u}\,du - x^{m+2}e^{-x}\int_0^x u^m e^{-u}\,du + (m+2)\frac{G_{2,m}(x)}{K(2, m)}\right]$$
$$\left. + 2\int_0^x u^{2m+3}e^{-2u}\,du\,\frac{G_{2,m+1}(x)}{K(2, m+1)} - x^{m+3}e^{-x}\frac{G_{3,m}(x)}{K(3, m)}\right\},$$

where

$$K(4, m) = 2^{4m+5}/[\Gamma(2m + 2)\Gamma(2m + 4)].$$

(d) $l = 5$.

$$\lim_{n\to\infty} \Pr (n\theta_1 \le x) = \lim_{n\to\infty} \Pr\left(\theta_1 \le \frac{x}{n}\right) = G_{5,m}(x)$$

$$= K(5, m)\left\{2\int_0^x u^{2m+7} e^{-2u}\, du \frac{G_{3,m}(x)}{k(3, m)} - 2\int_0^x u^{2m+6} e^{-2u}\, du\right.$$

$$\cdot\left[2\int_0^x u^{2m+4} e^{-2u}\, du \int_0^x u^m e^{-u}\, du - 2\int_0^x u^{2m+3} e^{-2u}\, du\right.$$

$$\left.\cdot\int_0^x u^{m+1} e^{-u}\, du - x^{m+3} e^{-x} \frac{G_{2,m}(x)}{K(2, m)} + (m + 3)\frac{G_{3,m}(x)}{K(3, m)}\right]$$

$$+ 2\int_0^x u^{2m+5} e^{-2u}\, du\left\{2\int_0^x u^{2m+5} e^{-2u}\, du \int_0^x u^m e^{-u}\, du\right.$$

$$- 2\int_0^x u^{2m+3} e^{-2u}\, du \int_0^x u^{m+2} e^{-u}\, du - x^{m+3} e^{-x}$$

$$\left.\left.\cdot\left[2\int_0^x u^{2m+2} e^{-2u}\, du - x^{m+2} e^{-x}\int_0^x u^m e^{-u}\, du + (m + 2)\frac{G_{2,m}(x)}{K(2, m)}\right]\right\}\right.$$

$$\left.- 2\frac{G_{3,m+1}(x)}{K(3, m + 1)}\int_0^x u^{2m+4} e^{-2u}\, du - x^{m+4} e^{-x}\frac{G_{4,m}(x)}{K(4, m)}\right\},$$

where

$$K(5, m) = 2^{4m+9}/[3\Gamma(m + 1)\Gamma(2m + 3)\Gamma(2m + 5)].$$

5. Limiting distribution of the smallest root. It was shown in [2] that the exact distribution of the smallest root can be obtained by using the relation

$$\Pr (\theta_l \le x) = 1 - \Pr (\theta_1 \le 1 - x \mid \nu, \mu)$$

This relation, however, does not help in obtaining the limiting distribution of the smallest root from that of the largest root. The limiting distribution of the smallest root can be obtained by the method illustrated below

(a) $l = 2$

The exact distribution of the smallest root θ_2 can be expressed as

$$\Pr (\theta_2 \le x) = c(2, m, n)\{(0, 2, 1, x; m, n) + (0, 2, x, 1, z; m, n)\},$$

where $z = 1$. Replacing x by x/n, we get

$$\Pr (\theta_2 \le x/n) = c(2, m, n)\{(0, 2, 1, x/n; m, n) + (0, 2, x/n, 1, z; m, n)\},$$

where

$$(0, 2, 1, x/n, m, n) = \frac{1}{m + n + 2}\left[2\int_0^{x/n} y^{2m+1}(1 - y)^{2n+1}\, dy\right.$$

$$\left.- (0, x/n; m + 1, n + 1)\int_0^{x/n} y^m(1 - y)^n\, dy\right],$$

and

$$(0, 2, x/n, 1, z; m, n) = \frac{1}{m+n+2}\Big[(0, x/n; m+1, n+1) \cdot \int_0^z y^m(1-y)^n\,dy - (0, z; m+1, n+1)\int_0^{x/n} y^m(1-y)^n\,dy\Big],$$

as obtained from (6) of [2].

The limiting distribution of θ_2 is

$$\text{(13)}\quad \Pr(\theta_2 \le x/n) = \lim_{n\to\infty} c(2, m, n)\{(0, 2, 1, x/n; m, n) + (0, 2, x/n, 1, z; m, n)\}.$$

Putting u/n for y, the variate of integration and allowing n to tend to infinity, we have

$$\lim_{n\to\infty} c(2, m, n)(0, 2, 1, x/n; m, n) = K(2, m)\left\{2\int_0^x u^{2m+1}e^{-2u}\,du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\,du\right\},$$

and

$$\lim_{n\to\infty} c(2, m, n)(0, 2, x/n, 1, z, m, n) = K(2, m)x^{m+1}e^{-x}\int_0^\infty u^m e^{-u}\,du = K(2, m)x^{m+1}e^{-x}\Gamma(m+1).$$

Substituting these results in (13) we have

$$\lim_{n\to\infty}\Pr(n\theta_2 \le x) = \lim_{n\to\infty}\Pr(\theta_2 \le x/n) = K(2, m)\left\{2\int_0^x u^{2m+1}e^{-2u}\,du - x^{m+1}e^{-x}\int_0^x u^m e^{-u}\,du + x^{m+1}e^{-x}\Gamma(m+1)\right\},$$

where

$$K(2, m) = 2^{2m+1}/[\Gamma(2m+2)].$$

(b) $l = 3$.

The exact distribution of the smallest root can be expressed as

$$\Pr(\theta_3 \le x) = c(3, m, n)[(0, 3, 2, 1, x; m, n) + (0, 3, 2, x, 1, z, m, n) + (0, 3, x, 2, 1, z; m, n)],$$

where $z = 1$.

Replacing x by x/n and allowing n to tend to infinity we have

$$\text{(14)}\quad \Pr(n\theta_3 \le x) = \lim_{n\to\infty} c(3, m, n)[(0, 3, 2, 1, x/n, m, n) + (0, 3, 2, x/n, 1, z, m, n) + (0, 3, x/n, 2, 1, z; m, n)]$$

The values of these components on the right hand side of the above equation are given below.

$$\lim_{n\to\infty} c(3, m, n)(0, 3, 2, 1, x/n; m, n) = G_{3,m}(x), \quad \text{given by} \quad (12),$$

$$\lim_{n\to\infty} c(3, m, n)(0, 3, 2, x/n, 1, z, m, n)$$

$$(15) \qquad \begin{aligned} &= K(3, m)\left\{\int_x^\infty u^m e^{-u}\,du \left[2\int_0^x u^{2m+3} e^{-2u}\,du \right.\right. \\ &\left. - x^{m+2} e^{-x} \int_0^x u^{m+1} e^{-u}\,du\right] - x^{m+2} e^{-x}\left[2\int_0^x u^{2m+1} e^{-2u}\,du \right. \\ &\left. - x^{m+1} e^{-x} \int_0^x u^m e^{-u}\,du\right] + x^{m+2} e^{-x}\int_x^\infty u^{m+1} e^{-u}\,du \int_0^x u^m e^{-u}\,du \\ &\left. - 2\int_x^\infty u^{m+1} e^{-u}\,du \int_0^x u^{2m+2} e^{-2u}\,du\right\}, \end{aligned}$$

and

$$\begin{aligned} \lim_{n\to\infty} c(3, m, n)(0, 3, x/n, 2, 1, z; m, n) &= K(3, m)\left\{\int_0^x u^m e^{-u}\,du \left[2\int_x^\infty u^{2m+3} e^{-2u}\,du \right.\right. \\ &\left. - x^{m+2} e^{-x}\int_x^\infty u^{m+1} e^{-u}\,du\right] - x^{m+2} e^{-x}\left[2\int_x^\infty u^{2m+1} e^{-2u}\,du - x^{m+1} e^{-x}\int_x^\infty u^m e^{-u}\,du\right] \\ &\left. + x^{m+2} e^{-x}\int_0^x u^{m+1} e^{-u}\,du \int_x^\infty u^m e^{-u}\,du - 2\int_0^x u^{m+1} e^{-u}\,du \int_x^\infty u^{2m+2} e^{-2u}\,du\right\}. \end{aligned}$$

Substituting in (14) we have,

$$\begin{aligned} \lim_{n\to\infty} \Pr(n\theta_3 \le x) &= \{2^{2m+3}/[\Gamma(m+1)\Gamma(2m+3)]\} \\ &\cdot\left\{2\int_0^\infty u^{2m+3} e^{-2u}\,du \int_0^x u^m e^{-u}\,du + 2\int_0^x u^{2m+3} e^{-2u}\,du \int_x^\infty u^m e^{-u}\,du \right. \\ &- 2\int_0^\infty u^{m+1} e^{-u}\,du \int_0^x u^{2m+2} e^{-2u}\,du - 2\int_0^x u^{m+1} e^{-u}\,du \int_x^\infty u^{2m+2} e^{-2u}\,du \\ &- 2x^{m+2} e^{-x}\int_0^\infty u^{2m+1} e^{-2u}\,du - 2x^{m+2} e^{-x}\int_0^x u^{2m+1} e^{-2u}\,du \\ &\left. + x^{2m+3} e^{-2x}\left(\int_0^x u^m e^{-u}\,du + \int_0^\infty u^m e^{-u}\,du\right)\right\} \end{aligned}$$

Or,

$$\begin{aligned} \lim_{n\to\infty} \Pr(n\theta_3 \le x) &= 2^{2m+3}/[\Gamma(m+1)\Gamma(2m+3)] \\ &\cdot\left\{\frac{\Gamma(2m+4)}{2^{2m+4}}\int_0^x u^m e^{-u}\,du + 2\int_0^x u^{2m+3} e^{-2u}\,du \int_x^\infty u^m e^{-u}\,du \right. \\ &- 2\Gamma(m+2)\int_0^x u^{2m+2} e^{-2u}\,du - 2\int_0^x u^{m+1} e^{-u}\,du \int_x^\infty u^{2m+2} e^{-2u}\,du \\ &- \frac{\Gamma(2m+2)}{2^{2m+1}} x^{m+2} e^{-x} - x^{m+2} e^{-x}\int_0^x u^{2m+1} e^{-2u}\,du + \Gamma(m+1)x^{2m+3} e^{-2x} \\ &\left. + x^{2m+3} e^{-2x}\int_0^x u^m e^{-u}\,du\right\}. \end{aligned}$$

Thus we have seen that this method can be used for obtaining the limiting distribution of the smallest root for any value of l

6. Limiting distribution of any intermediate root. The above method can also be used for obtaining the limiting distribution of any intermediate root. We shall give the distribution of θ_2 for $l = 3$ We have

(16) $\Pr(\theta_2 \leq x) = c(3, m, n)\{(0, 3, 2, 1, x; m, n) + (0, 3, 2, x, 1, z, m, n)\},$

where $z = 1$.

The $\lim_{n\to\infty} c(3, m, n)(0, 3, 2, 1, v/n, m, n)$ and $\lim_{n\to\infty} c(3, m, n)(0, 3, 2, v/n, 1, z; m, n)$ are given by (12) and (15) respectively. Substituting these results in (16) and simplifying we get

$$\lim_{n\to\infty} \Pr(n\theta_2 \leq x) = \frac{2^{2m+3}}{\Gamma(m+1)\Gamma(2m+3)}\Bigg\{2\int_0^\infty u^m e^{-u}\,du \cdot \int_0^x u^{2m+3}e^{-2u}\,du - 2\int_0^\infty u^{m+1}e^{-u}\,du \int_0^x u^{2m+2}e^{-2u}\,du$$
$$- 4x^{m+2}e^{-x}\int_0^x u^{2m+1}e^{-2u}\,du + 2x^{2m+3}e^{-2x}\int_0^x u^m e^{-u}\,du$$
$$+ x^{m+2}e^{-x}\left[\int_x^\infty u^{m+1}e^{-u}\,du \int_0^x u^m e^{-u}\,du - \int_x^\infty u^m e^{-u}\,du \int_0^x u^{m+1}e^{-u}\,du\right]\Bigg\},$$

Or,

$$\lim_{n\to\infty} \Pr(n\theta_2 \leq x) = \frac{2^{2m+3}}{\Gamma(m+1)\Gamma(2m+3)}\Bigg\{2\Gamma(m+1)\int_0^x u^{2m+3}e^{-2u}\,du$$
$$- 2\Gamma(m+2)\int_0^x u^{2m+2}e^{-2u}\,du - 4x^{m+2}e^{-x}\int_0^x u^{2m+1}e^{-2u}\,du + 2x^{2m+3}e^{-2x}$$
$$\cdot \int_0^x u^m e^{-u}\,du + x^{m+2}e^{-x}\left[\int_x^\infty u^{m+1}e^{-u}\,du \int_0^x u^m e^{-u}\,du\right.$$
$$\left. - \int_x^\infty u^m e^{-u}\,du \int_0^x u^{m+1}e^{-u}\,du\right]\Bigg\}.$$

Thus the limiting distribution of any intermediate root can be obtained by the above method

7. Further problems. The limiting distribution of the largest root is found to be very helpful in obtaining the distribution of the sum of roots when $m = 0$. This condition implies that when the results are applied to canonical correlations the numbers of variates in the two sets differ by unity. The distributions for the sum of roots have been derived under the above condition for $l = 2, 3$ and 4 and the results are being presented in the next paper of this series

8. Acknowledgements. I am extremely thankful to Drs. P. L. Hsu and Harold Hotelling for guidance and help in this research

REFERENCES

[1] P L Hsu, "On the distribution of roots of certain determinantal equations," *Annals of Eugenics*, Vol. 3 (1939), pp 250–258.

[2] D N Nanda, "Distribution of a root of a determinantal equation," *Annals of Math Stat.*, Vol 18 (1947)

[3] S. N. Roy, "The individual sampling distribution of the maximum, the minimum and any intermediate of the '*p*' statistics on the null hypothesis," *Sankhyā*, Vol. 8 (1943).

ON A SOURCE OF DOWNWARD BIAS IN THE ANALYSIS OF VARIANCE AND COVARIANCE

By William G. Madow

Institute of Statistics, University of North Carolina

1. Summary. It is shown that if, in the analysis of variance, the experiments are not in a state of statistical control due to variations in the true means, then the test will have a downward bias. The power function of the analysis of variance test is obtained when this downward bias is present.

2. Introduction. To introduce the discussion of this bias let us consider the generalized Student's hypothesis.

Let $y_1, \cdots, y_{kN}$ be normally and independently distributed with variance σ^2, and let the expected value of $y_{i\nu}$, be $a_{i\nu}$.[1] Then the generalized Student's hypothesis is

(Null hypothesis) $$a_{i\nu} = a$$

and the class of alternative hypotheses against which the null hypothesis is tested is

(Class A) $$a_{i\nu} = a_i .$$

From the statement of the null hypothesis and the alternatives of Class A it follows that both the null hypothesis and the alternatives of Class A require that

$$a_{i1} = \cdots = a_{iN} . \tag{1.1}$$

Since our experiments are rarely in such perfect statistical control that (1.1) holds whether or not the null hypothesis is true, it becomes reasonable to investigate the existing F test when instead of the alternatives to the null hypothesis being of Class A, they are simply Class B:

(Class B) Equation (1.1) is false for at least one value of i

Furthermore, for many practical purposes we would prefer to test the average null hypothesis:

(Average null hypothesis) $$\bar{a}_i = \bar{a},$$

where $N\bar{a}_i = a_{i1} + \cdots + a_{iN}$ and $k\bar{a} = \bar{a}_1 + \cdots + \bar{a}_k$, instead of the null hypothesis, the alternatives to the average null hypothesis being of Class C.

(Class C) The $a_{i\nu}$ can have any values such that not all the $\bar{a}_i$ equal $\bar{a}$.

[1] Throughout this paper the letter i will assume all integral values from 1 to k, the letters μ, ν will assume all integral values from 1 to N, the letters γ, η will assume all integral values from 1 to m, the letter α will assume all integral values from $n_1 + \cdots + n_{\gamma-1} + 1$ to $n_1 + \cdots + n_\gamma$, $(n_0 = 0)$, and α_1, α_2 will assume all integral values from 0 to ∞

The F-test of the null hypothesis against the alternatives of Class A is, as is well known,

$$F = \frac{k(N - 1) \sum_{i} (\bar{y}_i - \bar{y})^2}{(k - 1 \sum_{i,\nu} (y_{i\nu} - \bar{y}_i)^2)}$$

where $N\bar{y}_i = y_{i1} + \cdots + y_{iN}$ and $k\bar{y} = \bar{y}_1 + \cdots + \bar{y}_k$. To answer the questions formulated above concerning the F-test when the average null hypothesis or the alternatives of classes B or C are true, we must then calculate the distribution of F under these various conditions This is done in Section 3.

A somewhat informal means of obtaining the conclusions is that of studying F itself. Taking the expected values of the numerator and denominator of F and defining

$$\phi_1^2 = \frac{N \sum_{i} (\bar{a}_i - \bar{a})^2}{(k - 1)\sigma^2}$$

$$\phi_2^2 = \frac{1}{k(N - 1)\sigma^2} \sum_{i,\nu} (a_{i\nu} - \bar{a}_i)^2$$

we obtain as the ratio of the two expected values

$$\bar{F} = \frac{1 + \phi_1^2}{1 + \phi_2^2}.$$

It is well known that, in general, the larger the value of N the more closely will F approximate $\bar{F}$. From this fact it is easy to see why if the null hypothesis is true, then $F \sim 1$, whereas if the null hypothesis is false but an alternative of Class A is true then

$$F \sim 1 + \phi_1^2 > 1$$

so that large values of F become more likely than if the null hypothesis were true. However, if an alternative of Class B is true then

$$F \sim \frac{1 + \phi_1^2}{1 + \phi_2^2}$$

so that if $\phi_1^2 < \phi_2^2$, smaller values of F occur more frequently than indicated by the null hypothesis. Thus we would tend to accept the null hypothesis more frequently than desired when it is false. Even when the null hypothesis is false so that $\phi_1^2 > 0$, the values of F will tend to be less if $\phi_2^2 > 0$ than if $\phi_2^2 = 0$ whether or not $\phi_1^2 < \phi_2^2$. Not only is the probability of an error of the first kind less than the value ϵ we may have previously selected, but also the power of the test is less than would be indicated by Tang's tables [1]. The lack of statistical control represented by variation of expected values within a class has the effect of making it less likely than the standard F-test indicates that the null

hypothesis will be rejected whether it be true or false. Furthermore, even for relatively low values of ϕ_2^2, the reductions in the probabilities of rejection may be over 40 per cent as indicated by some examples given below.

If the average null hypothesis is true but (1 1) is false it follows that

$$F \sim \frac{1}{1 + \phi_2^2},$$

so that the full effect of the downward bias occurs in that case. Thus in cases where statistical control is lacking, to test the average null hypothesis by the F-test may well result in accepting the hypothesis when it is false. If the null hypothesis is rejected, however, then we can expect that the differences among the true means are even larger than indicated by Tang's tables.

To illustrate, it is shown in Section 4 that if $k = 5$ and $N = 7$, then the probability of rejecting the average null hypothesis when it is true, but (1.1) is false will not be the preassigned .05 but something less than .03 if $\phi_2^2 > .05$. Furthermore, if $\phi_2^2 > 07$, then the power of the F tests for this example will be reduced by at least 40 per cent whatever the value of ϕ_1^2.

The conclusions reached above remain valid for the analysis of variance and covariance in general. In the general case however, the value of the average null hypothesis in simplifying the analysis may be considerably reduced since the parameter ϕ_1^2 no longer vanishes when the average null hypothesis is true For example, if $Ey_\nu = \beta_\nu x_\nu$, and if the average null hypothesis is $\bar{\beta} = 0$, where $N\bar{\beta} = \beta_1 + \cdot + \beta_N$, then upon calculating

$$\phi_1^2 = \frac{\left(\sum_\nu \beta_\nu x_\nu^2\right)^2}{\sigma^2 \sum_\nu x_\nu^2}$$

we see that ϕ_1^2 will not vanish in general if $\bar{\beta}$ vanishes

Although as shown above the average null hypothesis may not have too great importance in the case of regression, yet if the "variance between treatments" is a function of arithmetic means of the random variables as in the "pure" analysis of variance the average null hypothesis may well be very useful Simple examples of this are provided by the randomized block, Latin square, and similar designs.

The distributions that we shall need are given in Section 3. The inequalities on the basis of which the bias is demonstrated are obtained in Section 4.

It would be highly desirable to have Tang's tables extended so that they might provide the answers to the questions raised by this source of bias In the absence of such extensions the inequalities of Section 4 may give some rough idea, but these inequalities are not sharp enough.

3. The calculation of the distributions. The following theorem was proved, although not explicitly stated, as part of an earlier note [2] (Note the change from x_i to y_i as the notation for the random variable.)

THEOREM 1. *Let $y_1, \cdots, y_N$ be normally and independently distributed with variance σ^2 and means $a_1, \cdots, a_N$ and let $q_1, \cdots, q_m$ be quadratic forms*

$$q_\gamma = \sum_{\mu,\nu} a_{\mu\nu}^{(\gamma)} y_\mu y_\nu$$

in $y_1, \cdots, y_N$ of ranks $n_1, \cdots, n_m$. Then, if an orthogonal transformation

$$y_\nu = \sum_\mu c_{\nu\mu} z_\mu$$

exists such that

$$q_\gamma = \sum_\alpha z_\alpha^2, \tag{2.1}$$

it follows that the random variables q_γ/σ^2 are independently distributed in χ'^2 distributions with degrees of freedom $n_1, \cdots, n_m$ and parameters $\lambda_1, \cdots, \lambda_m$, where

$$\lambda_\gamma = \frac{1}{2\sigma^2} \sum_{\mu,\nu} a_{\mu\nu}^{(\gamma)} a_\mu a_\nu = \frac{Eq_\gamma}{2\sigma^2} - \frac{n_\gamma}{2}.$$

Various conditions for the existence of an orthogonal transformation satisfying (2.1) of Theorem 1 have been given Among these are:

1. *Cochran's* [3] *condition.* If $\sum_\gamma q_\gamma = \sum_\nu y_\nu^2$ then a necessary and sufficient condition for the existence of an orthogonal transformation satisfying (2.1) is $\sum_\gamma n_\gamma = N$.

2. *Craig's* [4] *condition.* If A_γ denotes the matrix $(a_{\mu\nu}^{(\gamma)})$ then a necessary and sufficient condition for the existence of an orthogonal transformation satisfying (2.1) is $A_\gamma A_\eta = \delta_{\gamma\eta} A_\gamma$ where $\delta_{\gamma\eta}$ is the null matrix if $\gamma \neq \eta$ and the identity matrix if $\gamma = \eta$.

3. *Linear Hypothesis condition.* (Kolodziejczyk [5]) If λ be the likelihood ratio test of a linear hypothesis and if $E^2 = 1 - \lambda^{2/N}$, then $E^2 = q_1/(q_1 + q_2)$ and an orthogonal transformation exists satisfying (2.1) with $m = 2$.

To summarize some results obtained by Tang [1], let us state

THEOREM 2. *If $\chi_1'^2$ and $\chi_2'^2$ are independently distributed in distributions with n_1 and n_2 degrees of freedom and parameters λ_1 and λ_2, and if*

$$E^2 = \frac{\chi_1'^2}{\chi_1'^2 + \chi_2'^2},$$

then the probability density of E^2 is

$$\begin{aligned} p = p(E^2 \mid \lambda_1, \lambda_2, n_1, n_2) &= e^{-\lambda_1-\lambda_2}(E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1} \\ &\cdot \sum_{\alpha_1,\alpha_2} \frac{\lambda_1^{\alpha_1}\lambda_2^{\alpha_2}\Gamma\left(\frac{n_1+n_2}{2}+\alpha_1+\alpha_2\right)}{\alpha_1!\,\alpha_2!\,\Gamma\left(\frac{n_1}{2}+\alpha_1\right)\Gamma\left(\frac{n_2}{2}+\alpha_2\right)} (E^2)^{\alpha_1}(1 - E^2)^{\alpha_2} \end{aligned} \tag{2.2}$$

By assigning certain values to λ_1 and λ_2 we obtain the following special cases of (2 2)

$$p_1 = p(E^2 \mid \lambda_1, 0, n_1, n_2) = e^{-\lambda_1}(E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1}$$

$$\cdot \sum_{\alpha_1} \frac{\lambda_1^{\alpha_1}\Gamma\left(\frac{n_1 + n_2}{2} + \alpha_1\right)}{\alpha_1!\,\Gamma\left(\frac{n_1}{2} + \alpha_1\right)\Gamma\left(\frac{n_2}{2}\right)} (E^2)^{\alpha_1} \tag{2 3}$$

$$p_2 = p(E^2 \mid 0, \lambda_2, n_1, n_2) = e^{-\lambda_2}(E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1}$$

$$\cdot \sum_{\alpha_2} \frac{\lambda_2^{\alpha_2}\Gamma\left(\frac{n_1 + n_2}{2} + \alpha_2\right)}{\alpha_2!\,\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2} + \alpha_2\right)} (1 - E^2)^{\alpha_2} \tag{2.4}$$

$$p_0 = p(E^2 \mid 0, 0, n_1, n_2) = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} (E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1}. \tag{2.5}$$

It is noted that (2.3) is Tang's distribution (112) upon which the calculations of his tables were based. To see this we need only make the correspondence

This paper	*Tang*
λ_1	λ
n_1, n_2	f_1, f_2
α_1	i

We define ϵ to be the probability of an error of the first kind. Tang obtained the critical values E_ϵ^2 of E^2 by requiring that

$$P_I = \int_{E_\epsilon^2}^1 p_0 \, dE^2$$

$$= \epsilon \qquad \text{say } .01 \text{ or } .05.$$

Then he calculated

$$P_{II} = \int_0^{E_\epsilon^2} p_1(E^2 \mid \lambda_1, 0, n_1, n_2) \, dE^2$$

using the values of E_ϵ^2 obtained above. Hence $1 - P_{II}$ is the power of the test.

If, however, $\lambda_1 = 0$ but $\lambda_2 \neq 0$, then to find

$$P_{III} = \int_{E_\epsilon^2}^1 p_2(E^2 \mid 0, \lambda_2, n_1, n_2) \, dE^2$$

we could make the transformation $G^2 = 1 - E^2$ and find

$$P_{III} = \int_0^{1-E_\epsilon^2} p(G^2 \mid 0, \lambda_2, n_1, n_2)\, dG^2.$$

It is easy to verify that

$$p(G^2 \mid 0, \lambda_2, n_1, n_2) = p_1(E^2 \mid \lambda_2, 0, n_2, n_1)$$

if we put G in place of E^2 in the latter density. It follows that to calculate P_{III} it would be sufficient to have full tables of Tang's distribution since

$$P_{III} = \int_0^{1-E_\epsilon^2} p_1(E^2 \mid \lambda_2, 0, n_2, n_1)\, dE^2.$$

Tang's tables are not however sufficiently extensive. Furthermore, tables of (2 2) are also necessary. As yet these tables do not exist. However, some useful conclusions can be drawn from the inequalities obtained in the following section.

First, however, let us evaluate n_1, n_2, λ_1 and λ_2 for the generalized Student's hypothesis discussed in the introduction. It is easy to see that $n_1 = k - 1$ and $n_2 = k(N - 1)$. To evaluate λ_1 and λ_2 we note from Theorem 1 that we only need substitute Ey_{ij} for y_{ij} in q_1 and q_2 where

$$q_1 = N \sum_i (\bar{y}_i - \bar{y})^2$$

$$q_2 = \sum_{i,\nu} (y_{i\nu} - \bar{y}_i)^2.$$

Upon making these substitutions we obtain

$$\lambda_1 = \frac{N}{2\sigma^2} \sum_i (\bar{a}_i - \bar{a})^2$$

$$\lambda_2 = \frac{1}{2\sigma^2} \sum_{i,\nu} (a_{i\nu} - \bar{a}_i)^2.$$

Thus the various hypotheses concerning the a_{ij} influence the distribution of F or $E^2 = 1/(1 + Fn_1/n_2)$ by affecting the values of λ_1 and λ_2.

4. Limits of the values of p. It follows readily from (2.2) that,

$$p = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} (E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1} \cdot e^{-\lambda_1-\lambda_2} \sum_{\alpha_1,\alpha_2} \frac{\lambda_1^{\alpha_1}\lambda_2^{\alpha_2}}{\alpha_1!\,\alpha_2!} (E^2)^{\alpha_1}(1 - E^2)^{\alpha_2} C_{\alpha_1\alpha_2} \tag{3.1}$$

where

$$C_{\alpha_1\alpha_2} = \frac{\Gamma\left(\frac{n_1+n_2}{2} + \alpha_1 + \alpha_2\right)\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2} + \alpha_1\right)\Gamma\left(\frac{n_2}{2} + \alpha_2\right)\Gamma\left(\frac{n_1+n_2}{2}\right)}.$$

Now if $a > 0, b > 0$, and j is an integer > 1, we have

$$\left(1+\frac{a}{b}\right)\left(1+\frac{a}{b+2}\right) \cdots \left(1+\frac{a}{b+2(j-1)}\right) < \left(1+\frac{a}{b}\right)^j .$$

Hence, it follows that

$$1 \leq C_{\alpha_1\alpha_2} \leq \left(\frac{n_1+n_2}{n_1}\right)^{\alpha_1}\left(\frac{n_1+n_2+2\alpha_1}{n_2}\right)^{\alpha_2} .$$

Substituting we see that

$$p_0 e^{-\lambda_1-\lambda_2} e^{\lambda_1 E^2+\lambda_2(1-E^2)} \leq p \leq p_0 e^{-\lambda_1-\lambda_2}$$
$$\cdot \exp\left\{\lambda_1 E^2\left(\frac{n_1+n_2}{n_1}\right)\exp\left[\frac{2\lambda_2(1-E^2)}{n_2}\right]+\lambda_2(1-E^2)\left(\frac{n_1+n_2}{n_2}\right)\right\} \tag{3.2}$$

and

$$p_1 e^{-\lambda_2+\lambda_2(1-E^2)} < p < p_1 \exp\left[-\lambda_2+\lambda_2(1-E^2)\left(\frac{n_1+n_2}{n_2}\right)+2\frac{\lambda_2}{n_2}\right]. \tag{3.3}$$

Let $2n_i\phi_i^2 = \lambda_i,\ i = 1, 2$.

THEOREM 3 *Let* $\epsilon = \int_{E_\epsilon^2}^1 p_0\, dE^2$ *so that* ϵ *is the probability of an error of the first kind. Then, for all values of* ϕ_2^2

$$\epsilon > \int_{E_\epsilon^2}^1 p_2\, dE^2 \tag{3 4}$$

and if $E^2 > n_1/(n_1+n_2)$, *it follows that*

$$\epsilon > \epsilon \exp\{-2n_2\phi_2^2 + 2\phi_2^2(1-E_\epsilon^2)(n_1+n_2)\} > \int_{E_1^2}^1 p_2\, dE^2 > \epsilon e^{-n_2\phi_2^2} . \tag{3.5}$$

Furthermore, for all values of ϕ_2^2

$$\int_{E_\epsilon^2}^1 p_1\, dE^2 > \int_{E_\epsilon^2}^1 p\, dE^2, \tag{3.6}$$

and if $E^2 > (n_1+2)/(n_1+n_2)$, *it follows that*

$$\int_{E_\epsilon^2}^1 p_1\, dE_2 > \exp\{-2n_2\phi_2^2 + 2\phi_2^2(1-E_\epsilon^2)(n_1+n_2)\ 2\phi_2^2\} \int_{E_\epsilon^2}^1 p_1\, dE^2$$
$$> \int_{E_\epsilon^2}^1 p\, dE^2 > e^{-2n_2\phi_2^2}\int_{E_\epsilon^2}^1 p_1\, dE^2. \tag{3.7}$$

Finally, if γ *can assume the two values* 0 *and* 2, *it follows that if*

$$\phi_2^2 > \frac{-\log\delta}{2(E_\epsilon^2(n_1+n_2)-(n_1+\gamma))} > 0, \tag{3.8}$$

then if $\gamma = 0$,

$$\int_{E_\epsilon^2}^1 p_2\, dE^2 < \epsilon\delta \tag{3 9}$$

and if $\gamma = 2$

$$\int_{E_\epsilon^2}^1 p\, dE^2 < \delta \int_{E_\epsilon^2}^1 p_1\, dE^2. \tag{3.10}$$

PROOF. To prove (3.4) and (3.6) it is only necessary to follow Daly's [6] procedure.[2] Since

$$\exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\}$$

and

$$\exp\{- n_2\phi_2^2 E^2\}$$

are decreasing functions of E^2, and

$$\exp\{- 2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\} < 1$$

if

$$E^2 > \frac{n_1 + \gamma}{n_1 + n_2}$$

the inequalities (3.5) and (3.7) follow immediately from (3.2) and (3.3). Finally

$$\exp\{- 2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\} < \delta < 1$$

if (3.8) is true, so that (3 9) and (3.10) follow.

From (3 8), (3.9) and (3.10) we can calculate either a lower limit for the bias, if we know ϕ_2, or the upper limit that ϕ_2 can have if we wish the bias to be not greater than some given amount. Thus these limits do not answer the important question of what is a value ϕ_2 such that if $\phi_2 < \phi$ then the bias is less than $(1 - \delta)\epsilon$. They only provide a value ϕ' of ϕ_2 such that if $\phi_2 > \phi'$ then the bias is at least $(1 - \delta)\epsilon$.

If, for example, $\delta = 5$ and $n_1 = 1$ as in the case of Students' ratio, we have if $\gamma = 0$

$$\phi_2^2 > \frac{693}{2(n_2 E^2 - 1)}$$

and if $\epsilon = 05$, then E^2 decreases steadily from .903 if $n_2 = 2$, to .063 if $n_2 = 60$' and the corresponding lower limits of ϕ_2^2 decrease from .43 to .12. Thus, if $\phi_2^2 > 43$ or .12 in these two cases, it follows that the probability of rejecting the average null hypothesis will be not 05 but something less than .025.

If $\delta = .6$ and $n_1 = 4$, $n_2 = 30$ then we can evaluate the lower limit of ϕ_2^2 for the example given in the introduction finding.

$$\phi_2^2 > \frac{.511}{2(.279)(34) - 8} = .05$$

implies a downward bias of at least 40 per cent of .05. Also, if $\phi_2^2 > .07$ then for

[2] The procedure followed is given in [6] on pp. 4, 5, equations (2.2) through Lemma 1.

any value of ϕ_1 the power of the analysis of variance test is reduced at least 40 per cent.

5. Conclusions. The rather sharp effects of a moderate lack of statistical control on the probabilities associated with the F-test indicates the importance of testing for statistical control outside of the industrial applications now made. Furthermore, it would seem advisable to investigate tests and designs that are less sensitive to the lack of control than is the F-test

REFERENCES

[1] P C Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Statistical Research Memoirs*, Vol 2 (1938), pp. 126-157.

[2] William G Madow, "The distribution of quadratic forms in non-central normal random variables," *Annals of Mathem Stat*, Vol 9 (1940), pp 100-104.

[3] W G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Cambridge Phil. Soc Proc* Vol. 30 (1934), pp 178-191

[4] A. T. Craig, "Note on the independence of certain quadratic forms," *Annals of Math Stat.*, Vol 14 (1943), pp 195-197

[5] S. Kolodziejczyk, "On the important class of statistical hypotheses," *Biometrika*, Vol 27 (1935), pp 161-190.

[6] J. F. Daly, "On the unbiased character of likelihood-ratio tests for independence in normal systems," *Annals of Math Stat*, Vol. 11 (1940), pp 1-33. The procedure followed is given on pp 4, 5, equations (2 2) through Lemma 1

MIXTURE OF DISTRIBUTIONS

By Herbert Robbins

Department of Mathematical Statistics, University of North Carolina

1. Summary. Mixtures of measures or distributions occur frequently in the theory and applications of probability and statistics. In the simplest case it may, for example, be reasonable to assume that one is dealing with the mixture in given proportions of a finite number of normal populations with different means or variances. The mixture parameter may also be denumerably infinite, as in the theory of sums of a random number of random variables, or continuous, as in the compound Poisson distribution.

The operation of Lebesgue-Stieltjes integration, $\int f(x)\, d\mu$, is linear with respect to both integrand $f(x)$ and measure μ. The first type of linearity has as its continuous analog the theorem of Fubini on interchange of order of integration; the second type of linearity has a corresponding continuous analog which is of importance whenever one deals with mixtures of measures or distributions, and which forms the subject of the present paper. Other treatments of the same subject have been given ([1], [2]; see also [3], [4]) but it is hoped that the discussion given here will be useful to the mathematical statistician.

A general measure theoretic form of the fundamental theorem is given in Section 2, and in Section 3 the theorem is formulated in terms of finite dimensional spaces and distribution functions. The operation of convolution as an example of mixture is treated briefly in Section 4, while Section 5 is devoted to random sampling from a mixed population.

We shall refer to *Theory of the Integral* by S. Saks (second edition, Warszawa, 1937) as [*S*], and the *Mathematical Methods of Statistics* by H. Cramér (Princeton, 1946) as [*C*].

2. Mixture of measures in general. Let $X(Y)$ be a space with points $x(y)$ and let $\mathfrak{X}(\mathfrak{Y})$ be a σ-field of subsets of $X(Y)$. Let ν be a measure on $\mathfrak{Y}$. Let μ_y be for a.e. (ν) y a measure on $\mathfrak{X}$, such that $\mu_y(S)$ is for every S in $\mathfrak{X}$ a measurable $(\mathfrak{Y})$ function of y. Define for every S in $\mathfrak{X}$,

$$\mu(S) = \int_Y \mu_y(S)\, d\nu. \tag{1}$$

Theorem 1. *μ is a measure on $\mathfrak{X}$. If $\nu(Y) = \mu_y(X) = 1$, then $\mu(X) = 1$.*

Proof. Clear.

Theorem 2. *If $f(x)$ is any non-negative or non-positive function measurable $(\mathfrak{X})$ then the function*

$$g(y) = \int_X f(x)\, d\mu_y \tag{2}$$

is measurable ($\mathfrak{Y}$), and

$$\int_X f(x)\, d\mu = \int_Y g(y)\, d\nu. \tag{3}$$

PROOF. First let $f_0(x)$ be any non-negative simple function [S, p. 7] of the form

$$f_0(x) = \{a_1, S_1; \cdots; a_k, S_k\} \tag{4}$$

where the S_i are disjoint sets in $\mathfrak{X}$ such that $X = \sum_1^k S_i$ and the a_i are non-negative constants Then

$$g_0(y) = \int_X f_0(x)\, d\mu_y = \sum_1^k a_i\, \mu_y(S_i) \tag{5}$$

is a non-negative function measurable ($\mathfrak{Y}$), and from (1) it follows that each side of (3) is equal to $\sum_1^k a_i \mu(S_i)$ Hence the theorem holds in this case.

Next let $f(x)$ be any non-negative function measurable ($\mathfrak{X}$); then [S, p. 14] there exists a sequence $f_n(x)$ of simple functions such that for every x,

$$0 \le f_1(x) \le f_2(x) \le \cdots; \qquad \lim_{n\to\infty} f_n(x) = f(x). \tag{6}$$

Setting

$$g_n(y) = \int_X f_n(x)\, d\mu_y, \qquad g(y) = \int_X f(x)\, d\mu_y, \tag{7}$$

it follows from the theorem of monotone convergence [S, p. 28] and from the preceding paragraph that

$$\int_X f(x)\, d\mu = \lim_{n\to\infty} \int_X f_n(x)\, d\mu = \lim_{n\to\infty} \int_Y g_n(y)\, d\nu, \tag{8}$$

$$g(y) = \lim_{n\to\infty} \int_X f_n(x)\, d\mu_y = \lim_{n\to\infty} g_n(y). \tag{9}$$

From (6) and (9) it follows that for a.e. $(\nu)y$,

$$0 \le g_1(y) \le g_2(y) \le \cdots, \qquad \lim_{n\to\infty} g_n(y) = g(y). \tag{10}$$

Hence $g(y)$ is measurable ($\mathfrak{Y}$), and from the theorem of monotone convergence,

$$\int_Y g(y)\, d\nu = \lim_{n\to\infty} \int_Y g_n(y)\, d\nu. \tag{11}$$

Equation (3) now follows from (8) and (11).

By passing from $f(x)$ to $-f(x)$ we establish (3) when $f(x)$ is any non-positive function measurable ($\mathfrak{X}$). This completes the proof of Theorem 2

If $f(x)$ is an arbitrary function measurable ($\mathfrak{X}$) we define

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) \ge 0 \\ 0 & \text{otherwise} \end{cases}, \qquad f^-(x) = \begin{cases} f(x) & \text{if } f(x) \le 0 \\ 0 & \text{otherwise} \end{cases}, \tag{12}$$

so that

$$f(x) = f^+(x) + f^-(x) \tag{13}$$

is the sum of two functions measurable ($\mathfrak{X}$) of constant sign. By Theorem 2 the functions

$$g_1(y) = \int_X f^+(x)\, d\mu_y, \qquad g_2(y) = \int_X f^-(x)\, d\mu_y \tag{14}$$

are measurable ($\mathfrak{Y}$) and

$$0 \leq \int_X f^+(x)\, d\mu = \int_Y g_1(y)\, d\nu \leq \infty, \tag{15}$$

$$0 \geq \int_X f^-(x)\, d\mu = \int_Y g_2(y)\, d\nu \geq -\infty. \tag{16}$$

The integral $\int_X f(x)\, d\mu$ exists if and only if at least one of the two quantities (15) and (16) is finite [S, p 20].

THEOREM 3. *A necessary and sufficient condition that*

$$\int_X f(x)\, d\mu = \int_Y \left\{\int_X f(x)\, d\mu_y\right\} d\nu \tag{17}$$

is that at least one of the two quantities (15) *and* (16) *be finite.*

PROOF. By the remark preceding Theorem 3 the condition is clearly necessary. Now suppose, e.g., that (15) is finite; we must show that (17) holds. By hypothesis,

$$\int_X f^+(x)\, d\mu < \infty, \qquad \int_X f(x)\, d\mu = \int_X f^+(x)\, d\mu + \int_X f^-(x)\, d\mu. \tag{18}$$

From (18) and (15) it follows that $0 \leq g_1(y) < \infty$ for a.e. $(\nu)y$; hence

$$\int_X f(x)\, d\mu_y = \int_X f^+(x)\, d\mu_y + \int_X f^-(x)\, d\mu_y = g_1(y) + g_2(y) \tag{19}$$

exists for a.e. $(\nu)y$. From the finiteness of (15) it follows that

$$\int_Y (g_1(y) + g_2(y))\, d\nu = \int_Y g_1(y)\, d\nu + \int_Y g_2(y)\, d\nu \tag{20}$$

exists. Hence from (19), the integral

$$\int_Y \left\{\int_X f(x)\, d\mu_y\right\} d\nu = \int_Y (g_1(y) + g_2(y))\, d\nu \tag{21}$$

exists. Equation (17) now follows from (21), (20), (15), and (18). This completes the proof of Theorem 3

COROLLARY 1. *If $\mu(X) < \infty$, and if $f(x)$ is bounded from above or from below, then both sides of* (17) *exist and the equality holds.*

PROOF. If, say, $f(x) \leq C < \infty$, then

$$0 \leq \int_X f^+(x)\, d\mu \leq C \cdot \mu(X) < \infty,$$

and the result follows from Theorem 3

We shall now show by an example that the existence and even the finiteness of the right side of (17) does not imply the existence of the left side.

Let $X = Y = \{1, 2, \cdots, n, \cdots\}$ and let $\mathfrak{X}(\mathfrak{Y})$ consist of all subsets of $X(Y)$. Let ν be the measure which assigns mass c_n to n, where the c_n are positive constants such that $\sum_1^\infty c_n = 1$. Let μ_n assign the mass $1/2n$ to each of the points $1, 2, \cdots, 2n$. Let $f(x)$ be such that $f(1) = b_1, f(2) = -b_1, f(3) = b_2, f(4) = -b_2, \cdots$ where the b_n are positive constants. Then

$$\int_X f(x)\, d\mu_n = 0 \qquad (n = 1, 2, \cdots),$$

so that

$$\int_Y \left\{ \int_X f(x)\, d\mu_n \right\} d\nu = 0.$$

The measure μ defined by (1) assigns to each n a positive value $\mu(n)$ given by

$$\mu(1) = \mu(2) = c_1 \cdot (2)^{-1} + c_2 \cdot (2 \cdot 2)^{-1} + c_3 \cdot (2 \cdot 3)^{-1} + \cdots$$

$$\mu(3) = \mu(4) = c_2 \cdot (2 \cdot 2)^{-1} + c_3 \cdot (2 \cdot 3)^{-1} + \cdots$$

$$\cdots$$

where $\mu(X) = \sum_1^\infty \mu(n) = \sum_1^\infty c_n = 1$.

Now fix the b_n and c_n in such a way that

$$b_1 \cdot \mu(1) + b_2 \cdot \mu(3) + b_3 \cdot \mu(5) + \cdots = \infty.$$

Then

$$\int_X f^+(x)\, d\mu = -\int_X f^-(x)\, d\mu = \infty,$$

so that the left side of (17) does not exist, even though $\nu(Y) = \mu_y(X) = \mu(X) = 1$ and the right side of (17) exists and is equal to zero.

3. A restatement of the preceding results in the form most useful in probability theory. Let $x = (x_1, \cdots, x_n)$ be a point in the n-dimensional Euclidean space R_n, and let B_n denote the σ-field of Borel sets in R_n. Let S_x denote the half-open interval in R_n consisting of all points $(w_1, \cdots, w_n)$ in R_n satisfying the inequalities

$$w_1 \leq x_1, \cdots, w_n \leq x_n; \tag{22}$$

then if μ is any probability measure on B_n the function

$$F(x) = \mu(S_x) \tag{23}$$

is the distribution function corresponding to μ Conversely, if $F(x)$ is any distribution function in R_n [C, p. 80] there is a unique probability measure μ on B_n such that (23) holds. As a matter of notation we write for any Borel measurable $f(x)$,

$$\int_{R_n} f(x)\, d\mu = \int_{-\infty}^{\infty} f(x)\, dF(x) \tag{24}$$

provided the integral on the left exists.

Now let $y = (y_1, \cdots, y_m)$ be a point in R_m, let $G(y)$ be a distribution function, and let ν denote the corresponding probability measure on B_m. Let $F(x,y)$ be for a e $(\nu)y$ a distribution function in x, and for every x a Borel measurable function of y, and let μ_y be the corresponding probability measure on B_n.

THEOREM 4 *The function*

$$H(x) = \int_{-\infty}^{\infty} F(x, y)\, dG(y) \tag{25}$$

is a distribution function in R_n. Let μ denote the corresponding probability measure on B_n. Then for any S in B_n, $\mu_y(S)$ is a Borel measurable function of y and

$$\mu(S) = \int_{-\infty}^{\infty} \mu_y(S)\, dG(y). \tag{26}$$

PROOF. Let C denote the class of all Borel sets S in R_n such that $\mu_y(S)$ is a Borel measurable function of y. We shall show that C is a normal class [S, p 83].

(i) If $S_1, S_2, \cdots$ is a sequence of disjoint sets in C and if $S = \sum_1^{\infty} S_n$, then

$$\mu_y(S) = \mu_y\left(\sum_1^{\infty} S_n\right) = \sum_1^{\infty} \mu_y(S_n)$$

is a convergent series of Borel measurable functions and is therefore itself a Borel measurable function.

(ii) If $S_1 \supset S_2 \supset \cdots$ is a decreasing sequence of sets in C and if $S = \prod_1^{\infty} S_n$, then

$$\mu_y(S) = \mu_y\left(\prod_1^{\infty} S_n\right) = \lim_{n\to\infty} \mu_y(S_n)$$

is the limit of a sequence of Borel measurable functions and is therefore a Borel measurable function.

Hence C is a normal class But C contains every interval S_x, for $\mu_y(S_x) = F(x, y)$ was assumed to be a Borel measurable function of y for every x. It follows [S, p 85] that $C = B_n$

It now follows from Theorem 1 that the set function $\mu(S)$ defined by (26) is a probability measure on B_n. The corresponding distribution function is the function $H(x)$ defined by (25). Thus Theorem 4 is proved

Let $f(x) = f^+(x) + f^-(x)$ be any Borel measurable function. Then from Theorem 2, the integrals

(27)
$$\int_{-\infty}^{\infty} f^+(x)\, dH(x) = \int_{-\infty}^{\infty} f^+(x)\, d_x\left\{\int_{-\infty}^{\infty} F(x, y)\, dG(y)\right\} = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} f^+(x)\, d_x F(x, y)\right\} dG(y),$$

(28)
$$\int_{-\infty}^{\infty} f^-(x)\, dH(x) = \int_{-\infty}^{\infty} f^-(x)\, d_x\left\{\int_{-\infty}^{\infty} F(x, y)\, dG(y)\right\} = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} f^-(x)\, d_x F(x, y)\right\} dG(y)$$

exist. The following theorem is an immediate consequence of Theorem 3 and Corollary 1

THEOREM 5. *A necessary and sufficient condition that*

(29)
$$\int_{-\infty}^{\infty} f(x)\, d_x\left\{\int_{-\infty}^{\infty} F(x, y)\, dG(y)\right\} = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} f(x)\, d_x F(x, y)\right\} dG(y)$$

is that the left side of (29) *exist; i e that at least one of the quantities* (27) *and* (28) *be finite. This will be true in particular if* $f(x)$ *is bounded from above or from below*

4. The operation of convolution. An example of the general mixture (25) of distribution functions is the operation of convolution: if $F(x)$, $G(x)$ are two distribution functions in R_1 then $F(x, y) = F(x - y)$ satisfies the conditions of Theorem 4, so that

(30)
$$H(x) = \int_{-\infty}^{\infty} F(x - y)\, dG(y)$$

is also a distribution function in R_1, denoted by

(31)
$$H(x) = F(x) * G(x).$$

Corresponding to any distribution function $F(x)$ in R_1 is the characteristic function

(32)
$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)$$

which in turn uniquely determines $F(x)$ [C, p. 93].

THEOREM 6. *Let* $F(x)$, $G(x)$, $H(x)$ *be distribution functions in* R_1 *and let* $\varphi_1(t)$, $\varphi_2(t)$, $\varphi(t)$ *be the corresponding characteristic functions. Then*

(33)
$$H(x) = F(x) * G(x)$$

if and only if

(34)
$$\varphi(t) = \varphi_1(t) \cdot \varphi_2(t).$$

PROOF. Assume (33) holds. Since $|e^{itx}| \leq 1$ we have from Theorem 5,

$$
(35)\quad
\begin{aligned}
\varphi(t) &= \int_{-\infty}^{\infty} e^{itx}\, d_x\left\{\int_{-\infty}^{\infty} F(x-y)\, dG(y)\right\} \\
&= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} e^{itx}\, d_x F(x-y)\right\} dG(y) \\
&= \int_{-\infty}^{\infty} e^{ity}\left\{\int_{-\infty}^{\infty} e^{it(x-y)}\, d_x F(x-y)\right\} dG(y) \\
&= \int_{-\infty}^{\infty} e^{ity}\left\{\int_{-\infty}^{\infty} e^{itw}\, dF(w)\right\} dG(y) = \varphi_1(t)\cdot\varphi_2(t).
\end{aligned}
$$

The converse implication now follows from the fact that the characteristic function of a distribution determines the latter uniquely.

The importance of the operation $\bar{*}$ in probability theory arises from the fact that if X, Y are *independent* random variables with respective distribution functions $F(x)$, $G(x)$, and if $Z = X + Y$, then the distribution function $H(x)$ of Z satisfies (33), since for any value of a,

$$
(36)\quad
\begin{aligned}
H(a) &= P[X + Y \leq a] = \iint_{x+y\leq a} dF(x)\, dG(y) \\
&= \int_{-\infty}^{\infty}\left\{\int_{x\leq a-y} dF(x)\right\} dG(y) = \int_{-\infty}^{\infty} F(a-y)\, dG(y) = F(a) * G(a),
\end{aligned}
$$

the evaluation of the double integral by an iterated integral following from Fubini's theorem [S, pp. 76–88] However, (33) may hold without X, Y being independent, and Theorem 6 shows that (34) will then hold also, and conversely

An example where $H(x) = F(x) \dot{*} G(x)$ without X, Y being independent has been given by Cramér [C, p. 317, exercise 2] We shall give another Let points $0, A, \cdots, F$ in the (x, y)-plane be defined as follows.

$$
O = (0,0),\ A = (1,1),\ B = (1/2, 1),\ C = (0, 1/2),\ D = (1, 0),
$$
$$
E = (1, 1/2),\ F = (1/2, 0).
$$

Let $f(x, y)$ have the value 2 inside the quadrilateral $OABC$ and the triangle DEF, and 0 elsewhere. Then if $f(x, y)$ is the joint frequency function of X, Y it is easily seen that X and Y have uniform distributions on the intervals $0 \leq x \leq 1$, $0 \leq y \leq 1$ respectively and that $Z = X + Y$ has the triangular distribution given by (33), although X and Y are not independent.

It would be interesting to know what distribution functions $F(x)$ are such that if X, Y, $Z = X + Y$ are random variables with the distribution functions $F(x)$, $F(x)$, $F(x) * F(x)$ respectively, then X and Y are necessarily independent. A rather trivial example of such a distribution function is the step function $F(x)$ with jumps of $\frac{1}{2}$ at the points $x = 0$ and $x = 1$. It can be shown (oral communication by W. Hoeffding), in generalization of Cramér's example, that no abso-

lutely continuous distribution function (e.g. the normal distribution function) has this property.

5. The problem of random sampling from a mixed population. Let $G(v)$ be a distribution function in the real variable v, and let $F(u, v)$ be for a.e. (relative to the measure corresponding to G) v a distribution function in the real variable u, and for every u a Borel measurable function of v. Let

$$H(u) = \int_{-\infty}^{\infty} F(u, v)\, dG(v); \tag{37}$$

then by Theorem 4 $H(u)$ is a distribution function in R_1. Now define for $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$

$$\begin{aligned} \bar{H}(x) &= H(x_1) \cdots H(x_n), \\ \bar{G}(y) &= G(y_1) \cdots G(y_n). \end{aligned} \tag{38}$$

Both $\bar{H}(x)$ and $\bar{G}(y)$ are then distribution functions in R_n. In particular, $\bar{H}(x)$ is the distribution function of a random sample of n independent variates each with the distribution function (37) Set

$$\bar{F}(x, y) = F(x_1, y_1) \cdots F(x_n, y_n); \tag{39}$$

then for a. e. (relative to the measure corresponding to $\bar{G}$) y, $\bar{F}(x, y)$ is a distribution function in x, and for every x, $\bar{F}(x, y)$ is a Borel measurable function of y. By Fubini's theorem we have

$$\begin{aligned} \bar{H}(x) &= \int_{-\infty}^{\infty} F(x_1, y_1)\, dG(y_1) \cdot \cdots \cdot \int_{-\infty}^{\infty} F(x_n, y_n)\, dG(y_n) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(x_1, y_1) \cdots F(x_n, y_n)\, dG(y_1) \cdots dG(y_n) \\ &= \int_{-\infty}^{\infty} \bar{F}(x, y)\, d\bar{G}(y). \end{aligned} \tag{40}$$

Thus $\bar{H}(x)$ is itself a mixture in the sense of Theorem 4. It follows from Theorem 5 that for any Borel measurable function $f(x)$,

$$\int_{-\infty}^{\infty} f(x)\, d\bar{H}(x) = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} f(x)\, d_x \bar{F}(x, y) \right\} d\bar{G}(y), \tag{41}$$

if and only if the left side of (41) exists. When written out in full (41) becomes

$$\begin{aligned} &\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n)\, d_{x_1} \left\{ \int_{-\infty}^{\infty} F(x_1, y_1)\, dG(y_1) \right\} \\ &\qquad \cdots d_{x_n} \left\{ \int_{-\infty}^{\infty} F(x_n, y_n)\, dG(y_n) \right\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \right. \\ &\qquad \left. \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n)\, d_{x_1} F(x_1, y_1) \cdots d_{x_n} F(x_n, y_n) \right\} dG(y_1) \cdots dG(y_n). \end{aligned} \tag{42}$$

Equation (41) is of particular interest in connection with the distribution of a statistic $t = t(x_1, \cdots, x_n) = t(x)$. For any distribution function $J(x)$ let $K(t \mid J)$ denote the distribution function of t when x has the distribution function $J(x)$. If we set

$$f(x) = \begin{cases} 1 & \text{if } t(x) \leq t, \\ 0 & \text{otherwise,} \end{cases} \tag{43}$$

then

$$K(t \mid J) = \int_{-\infty}^{\infty} f(x)\, dJ(x). \tag{44}$$

Hence from (41),

$$\begin{aligned} K(t \mid H(x_1) \cdots H(x_n)) &= K(t \mid \bar{H}) = \int_{-\infty}^{\infty} K(t \mid \bar{F}(x, y))\, d\bar{G}(y) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} K(t \mid F(x_1, y_1) \cdots F(x_n, y_n))\, dG(y_1) \cdots dG(y_n). \end{aligned} \tag{45}$$

As an example, let $t(x)$ be Student's ratio

$$t = n^{\frac{1}{2}} \cdot \bar{x}/s, \tag{46}$$

let

$$F(u, v) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{1}{2}(y-v)^2}\, dy, \tag{47}$$

and let

$$G(v) = \begin{cases} 0 \text{ for } & v < -a, \\ \frac{1}{2} \text{ for } & -a \leq v < a, \\ 1 \text{ for } & a \leq v. \end{cases} \tag{48}$$

Then $H(u)$ will be the distribution function of a mixture in equal proportions of two normal populations with unit variances and with means $-a$, a respectively, and $K(t \mid H(x_1) \cdots H(x_n))$ will be the distribution function of t in random samples of n from this non-normal population. On the other hand, $K(t \mid F(x_1, y_1) \cdots F(x_n, y_n))$ will be the distribution function of t in sampling from successive normal populations with unit variances and means $y_1, \cdots, y_n$ respectively. Relation (45) now becomes

$$K(t \mid H(x_1) \cdots H(x_n)) = \sum_{y_1, \cdots, y_n} K(t \mid F(x_1, y_1) \cdots F(x_n, y_n))/2^n, \tag{49}$$

where the summation is over all 2^n sets $(y_1, \cdots, y_n)$, each y_i being either $-a$ or a Due to the complexity of $K(t \mid F(x_1, y_1) \cdots F(x_n, y_n))$ (the frequency function of which is discussed in a forthcoming paper by the author), relation

(49) is not very useful. In other cases (45) may afford a considerable simplification in the evaluation of the distribution function of a statistic obtained in random sampling from a mixed population.

REFERENCES

[1] W Feller, "On the integro-differential equations of purely discontinuous Markoff processes," *Am Math. Soc. Trans.*, Vol 48 (1940), p 488.

[2] R. H. Cameron and W. T. Martin, "An unsymmetric Fubini theorem," *Am Math. Soc. Bull.*, Vol 47 (1941), p 121.

[3] P R. Halmos, "The decomposition of measures," *Duke Math. Jour.*, Vol. 8 (1941), p. 386

[4] W. Feller, "On a general class of 'contagious' distributions," *Annals of Math. Stat.*, Vol. 14 (1943), p 389.

SOME APPLICATIONS OF THE MELLIN TRANSFORM IN STATISTICS

BY BENJAMIN EPSTEIN

Coal Research Laboratory, Carnegie Institute of Technology

1. Summary. It is well known that the Fourier transform is a powerful analytical tool in studying the distribution of sums of independent random variables. In this paper it is pointed out that the Mellin transform is a natural analytical tool to use in studying the distribution of products and quotients of independent random variables. Formulae are given for determining the probability density functions of the product and the quotient $\frac{\xi}{\eta}$, where ξ and η are independent positive random variables with p.d.f.'s $f(x)$ and $g(y)$, in terms of the Mellin transforms $F(s) = \int_0^\infty f(x)\, x^{s-1}\, dx$ and $G(s) = \int_0^\infty g(y) y^{s-1}\, dy$. An extension of the transform technique to random variables which are not everywhere positive is given A number of examples including Student's t-distribution and Snedecor's F-distribution are worked out by the technique of this paper.

2. Introduction. It is well known [2], [3] that the Fourier transform is a useful analytical tool for studying the distribution of the sums of independent random variables. It is our purpose in this paper to study another transform which is useful in studying the distribution of the product of independent random variables. While it is perfectly true that one can reduce the study of the distribution of the random variable $\xi = \xi_1 \xi_2 \cdots \xi_n$, the product of n independent random variables $\xi_1, \xi_2, \cdots, \xi_n$, to the study of the distribution of the random variable $\eta = \log \xi = \log \xi_1 + \log \xi_2 + \cdots + \log \xi_n$, the sum of n independent random variables, it seems worth while to study the distribution problem directly. There are advantages inherent in the direct attack on the distribution problem which are lost to a considerable degree, if the problem is so transformed that the Fourier transform becomes applicable. In this paper we shall show that the direct application of the Mellin transform to the study of the distribution of products of independent random variables yields results of interest.

3. Connection between Mellin transforms and products of independent random variables. The key reason for the importance of Fourier transforms in studying the distribution of sums of independent random variables depends on the following result: if ξ_1 and ξ_2 are independent random variables with continuous[1] probability density functions, (henceforth abbreviated as p d.f), $f_1(x)$ and $f_2(x)$, respectively, then the p d.f. $f(x)$ of the random variable $\xi = \xi_1 + \xi_2$ is expressible[1] as

$$f(x) = \int_{-\infty}^{\infty} f_1(x - y) f_2(y)\, dy = \int_{-\infty}^{\infty} f_2(x - y) f_1(y)\, dy. \tag{1}$$

[1] In this paper we shall assume throughout that we are dealing with random variables with continuous p d.f.'s The argument can be extended with some changes to distribution functions which are perfectly general, but for simplicity this will not be done here.

But since these expressions are just the Fourier convolutions of $f_1(x)$ and $f_2(x)$, it is small wonder that the Fourier transform plays such a basic role in studying the distribution properties of sums of independent random variables

Consider now the following result for products of independent random variables (4), (5): if ξ_1 is a random variable with continuous p d f $f_1(x)$ and ξ_2, independent of ξ_1, is a positive random variable with continuous p.d f $f_2(x)$, then the p d.f $f(x)$ of the random variable $\xi = \xi_1\xi_2$ is expressible[2] as

$$f(x) = \int_0^\infty \frac{1}{y} f_1\left(\frac{x}{y}\right) f_2(y)\, dy. \tag{2}$$

But equation (2) is precisely in the form of a Mellin convolution of $f_1(x)$ and $f_2(x)$ and therefore it may be expected that the Mellin transform should be useful in studying the distribution of products of independent random variables

It is useful to indicate briefly the properties of the Mellin transform A detailed treatment of this transform will be found in [6] and we shall, therefore, stress only those portions of the theory of Mellin transforms which are of importance in the field of statistics. By definition, the Mellin transform $F(s)$, corresponding to a function $f(x)$ defined only[3] for $x \geq 0$, is

$$F(s) = \int_0^\infty f(x)x^{s-1}\, dx \tag{3}$$

Under certain restrictions on $f(x)$ [6, p. 47], $F(s)$ considered as a function of the complex variable s is a function of exponential type, analytic in a strip parallel to the imaginary axis. The width of the strip is governed by the order of magnitude of $f(x)$ in the neighborhood of the origin and for large values of x and, in particular, the strip of analyticity becomes a half-plane if $f(x)$ decays exponentially as $x \to \infty$ There is a reciprocal formula enabling one to go from the transform $F(s)$ to the function $f(x)$. This transformation is:

$$f(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^{-s} F(s)\, ds \tag{4}$$

for all x where $f(x)$ is continuous and where the path of integration is any line parallel to the imaginary axis and lying within the strip of analyticity of $F(s)$.

---

[2] More generally [4, p 411], if ξ_1 and ξ_2 are independent random variables with continuous p d f.'s $f_1(x)$ and $f_2(x)$, then the p.d f. of the random variable $\xi = \xi_1\xi_2$ is expressible as

$$f(x) = \int_{-\infty}^\infty \frac{1}{|y|} f_1\left(\frac{x}{y}\right) f_2(y)\, dy = \int_{-\infty}^\infty \frac{1}{|y|} f_2\left(\frac{x}{y}\right) f_1(y)\, dy. \tag{2'}$$

In [4] analogous results are given for random variables with perfectly general distribution functions.

[3] The reason for this restriction is that there are technical difficulties in defining a Mellin transform directly for a function defined over $(-\infty, \infty)$ In [6], for instance, the Mellin transform theory is given for functions defined only for positive values of the argument In statistical terminology this means that we are restricting ourselves for the moment to positive random variables This is, of course, an unnatural restriction and we shall indicate later in the paper a simple device for treating such questions

If, in particular, we are interested in applying Mellin transforms to p.d.f.'s of positive[4] random variables, the analysis can be carried out rigorously. Also, as in the case of the Fourier transform, one has the desirable property that there is a one-one correspondence between p.d.f.'s and their transforms.

A number of common p.d.f.'s of positive random variables have simple Mellin transforms. For example see Table 1.

In terms familiar to the mathematical statistician, the Mellin transform of a positive random variable ξ with continuous p.d.f. $f(x)$ is $E(\xi^{s-1})$, where

$$F(s) = E(\xi^{s-1}) = \int_0^\infty x^{s-1} f(x)\, dx \tag{5}$$

The following three basic properties hold: (*i*) The positive random variable $\eta = a\,\xi$, $a > 0$ has the Mellin transform $G(s) = a^{s-1}\,F(s)$. This is immediate since

$$G(s) = E(\eta^{s-1}) = E(a^{s-1}\,\xi^{s-1}) = a^{s-1}\,F(s). \tag{6}$$

(ii) The positive random variable $\eta = \xi^\alpha$ has the Mellin transform $G(s) = F(\alpha s - \alpha + 1)$. To prove this we note that

$$G(s) = E(\eta^{s-1}) = E(\xi^{\alpha s-\alpha}) = F(\alpha s - \alpha + 1). \tag{7}$$

In particular if $\alpha = -1$, i.e., $\eta = \dfrac{1}{\xi}$, then

$$G(s) = F(-s + 2).$$

This is a result which we shall have occasion to use later in the paper.

(iii) If ξ_1 and ξ_2 are independent positive random variables with Mellin transforms $F_1(s)$ and $F_2(s)$, respectively, then the Mellin transform of the product $\eta = \xi_1\xi_2$ is $G(s) = F_1(s)\,F_2(s)$. This is immediate since

$$\begin{aligned} G(s) = E(\eta^{s-1}) = E[(\xi_1\xi_2)^{s-1}] &= E(\xi_1^{s-1})\,E(\xi_2^{s-1}) \\ &= F_1(s)\,F_2(s). \end{aligned} \tag{8}$$

More generally if $\xi_1, \xi_2, \cdots, \xi_n$ are independent positive random variables with Mellin transforms $F_1(s), F_2(s), \cdots, F_n(s)$, then the Mellin transform of the random variable $\eta = \xi_1\xi_2 \cdots \xi_n$ is $G(s) = F_1(s)\,F_2(s) \cdots F_n(s)$. This relationship is fundamental and justifies the introduction of Mellin transforms in studying products of independent random variables.

From (8) it is clear that we can find the p.d.f. $g(y)$ of the random variable η which is the product of two positive independent random variables ξ_1 and ξ_2 with continuous p.d.f.'s $f_1(x)$ and $f_2(x)$. In fact, by the Mellin inversion formula

$$g(y) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} y^{-s}\,G(s)\,ds = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} y^{-s}\,F_1(s)F_2(s)\,ds, \tag{9}$$

[4] See footnote 3.

TABLE 1

	p d f	Mellin Transform	Region of Analyticity of Transform
(a)	$f(x) = 1, \quad 0 \le x \le 1$ $= 0$, elsewhere	$F(s) = \frac{1}{s}$	Half-plane, Re $(s) > 0$
(b)	$f(x) = \frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)}, 0 < x < \infty$ $\alpha > -1$	$F(s) = \frac{\Gamma(\alpha+s)}{\Gamma(\alpha+1)}$	Half-plane, Re $(s) > -\alpha$
(c)	$f(x) = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)} x^\alpha(1-x)^\beta, \quad 0 < x < 1$ $= 0$, elsewhere $\alpha > -1, \beta > -1$	$F(s) = \frac{\Gamma(\alpha+\beta+2)\Gamma(\alpha+s)}{\Gamma(\alpha+\beta+s+1)\Gamma(\alpha+1)}$	Half-plane, Re $(s) > -\alpha$
(d)	$f(x) = \frac{\Gamma(\beta)}{\Gamma(\alpha+1)\Gamma(\beta-\alpha-1)} \frac{x^\alpha}{(1+x)^\beta}, \quad 0 < x < \infty$ $\alpha > -1, \beta - \alpha > 1$	$F(s) = \frac{\Gamma(\alpha+s)\Gamma(\beta-\alpha-s)}{\Gamma(\alpha+1)\Gamma(\beta-\alpha-1)}$	Strip, $-\alpha <$ Re $(s) < \beta - \alpha$

where the path of integration is any line parallel to the imaginary axis and lying within the strip of analyticity of $G(s)$. As in the case of characteristic functions, it can be shown that there is a one-one correspondence between p.d.f.'s and their Mellin transforms Therefore, it follows that the p d.f. $g(y)$ computed in this way must be precisely equal to

$$(10) \qquad g(y) = \int_0^\infty \frac{1}{x} f_1\left(\frac{y}{x}\right) f_2(x)\, dx = \int_0^\infty \frac{1}{x} f_2\left(\frac{y}{x}\right) f_1(x)\, dx.$$

It is easy to verify this directly by showing that the Mellin transform of the right-hand side of (10) is $F_1(s)\, F_2(s)$ [6, p. 52], but this will not be done here. The essential point is that Equation (9), (which is sometimes easier to evaluate than Equation (10)), is a consequence of an algebraic formalism which is capable of revealing relationships which would otherwise remain hidden.

The p.d.f. $h(y)$ of $\eta = \frac{\xi_1}{\xi_2}$, the ratio of two positive random variables with continuous p.d.f 's, can be reduced to finding the p.d.f. of the product of independent random variables ξ_1 and $\frac{1}{\xi_2}$. If $F_1(s)$ and $F_2(s)$ are the Mellin transform corresponding to ξ_1 and ξ_2, respectively, then by (ii) $F_2(-s + 2)$ is the Mellin transform of $\frac{1}{\xi_2}$ and, therefore, the Mellin transform $H(s)$ of $\eta = \frac{\xi_1}{\xi_2}$ is $F_1(s)\, F_2(-s + 2)$. Therefore, the p.d.f. $h(y)$ of η is

$$(11) \quad h(y) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-s} H(s)\, ds = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-s} F_1(s) F_2(-s + 2)\, ds.$$

This formula is useful in finding distributions such as Student's t and Fisher's z.

4. A modified Mellin transform procedure for finding the distribution of the product of independent random variables which are not everywhere positive. Up to this point we have limited ourselves to the application of the Mellin transform to finding the distribution of the product or ratio of two positive independent random variables. While it is true that a number of interesting probability density functions are defined only for positive[5] values of the argument, it is certainly desirable that we be able to treat situations involving random variables capable of taking on both positive and negative values. A simple device for extending the Mellin transform treatment to the more general problem is to decompose the p.d f.'s $f_1(x)$ and $f_2(x)$ of the independent random variables ξ_1 and ξ_2 into

$$f_1(x) = f_{11}(x) + f_{12}(x),$$
$$f_2(x) = f_{21}(x) + f_{22}(x),$$

[5] For example, distributions of type 3, the χ^2 distribution, the distribution of the sample standard deviation and sample variance, the distribution of an even power of a random variable, etc are all defined only for positive values of the argument

where[6]

$$f_{11}(x) = 0, x < 0, \qquad f_{12}(x) = 0, x > 0,$$
$$f_{21}(x) = 0, x < 0, \qquad f_{22}(x) = 0, x > 0,$$

and then to operate on the pairs $[f_{11}(x), f_{21}(x)]$, $[f_{11}(x), f_{22}(x)]$, $[f_{12}(x), f_{21}(x)]$, and $[f_{12}(x), f_{22}(x)]$ separately. More specifically, the frequency distribution $h(y)$ corresponding to the random variable $\eta = \xi_1\xi_2$ is made up of the sum of four components $h_1(y)$, $h_2(y)$, $h_3(y)$, and $h_4(y)$. To compute $h_1(y)$ one can apply the Mellin transform directly to the evaluation of the expression

$$h_1(y) = \int_0^\infty \frac{1}{x} f_{11}\left(\frac{y}{x}\right) f_{21}(x)\, dx,$$

since both $f_{11}(x)$ and $f_{21}(x)$ are zero for negative values of x. The function $h_1(y)$ is zero for $y < 0$. To compute $h_2(y)$ we first evaluate

$$h_2^*(y) = \int_0^\infty \frac{1}{x} f_{11}\left(\frac{y}{x}\right) f_{22}(-x)\, dx.$$

Again $f_{11}(x)$ and $f_{22}(-x)$ are zero for negative values of x and, therefore, the conventional Mellin transform can be applied in determining $h_2^*(y)$. It is clear that $h_2^*(y) = 0$ for $y < 0$ and, therefore, $h_2(y) = h_2^*(-y) = 0$ for $y > 0$. Similarly, one can find $h_3(y)$ and $h_4(y)$ where $h_3(y) = 0$ for $y > 0$ and $h_4(y) = 0$ for $y < 0$, and it is readily seen that[7]

$$h(y) = h_1(y) + h_2(y) + h_3(y) + h_4(y)$$

is the desired p.d.f. of $\eta = \xi_1\xi_2$.

5. Examples of use of Mellin transforms in evaluating the product and quotient of independent random variables. Example 1: *The distribution of $\eta = \xi_1\xi_2$, where ξ_1 and ξ_2 are independent random variables with p.d.f.'s $f_1(x)$ and $f_2(x)$, respectively, where*

$$f_1(x) = f_2(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

In this case

$$f_1(x) = f_{11}(x) + f_{12}(x),$$

and

$$f_2(x) = f_{21}(x) + f_{22}(x),$$

[6] Of course, f_{11}, f_{12}, f_{21}, and f_{22} are generally not p.d.f.'s since $\int_0^\infty f_{11}(x)\, dx$, $\int_{-\infty}^0 f_{12}(x)\, dx$, $\int_0^\infty f_{21}(x)\, dx$, $\int_{-\infty}^0 f_{22}(x)\, dx$ are no longer necessarily equal to one.

[7] As in footnote 6, h_1, h_2, h_3, and h_4 are, in general, not p.d.f.'s.

where

$$f_{11}(x) = 0, x < 0; f_{12}(x) = 0, x > 0;$$
$$f_{21}(x) = 0, x < 0; f_{22}(x) = 0, x > 0.$$

The random variable $\eta = \xi_1\xi_2$ has a p.d.f. $h(y) = h_1(y) + h_2(y) + h_3(y) + h_4(y)$ where

$h_1(y)$ is associated with $[f_{11}(x), f_{21}(x)]$,
$h_2(y)$ is associated with $[f_{11}(x), f_{22}(x)]$,
$h_3(y)$ is associated with $[f_{12}(x), f_{21}(x)]$,
and $h_4(y)$ is associated with $[f_{12}(x), f_{22}(x)]$

It is sufficient to evaluate

$$h_1(y) = \int_0^\infty \frac{1}{x} f_{11}\left(\frac{y}{x}\right) f_{21}(x)\, dx.$$
$$= \int_0^\infty \frac{1}{x} f_{21}\left(\frac{y}{x}\right) f_{11}(x)\, dx.$$

In this case

$$F_{11}(s) = \int_0^\infty x^{s-1} f_{11}(x)\, dx = \int_0^\infty x^{s-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{2^{\frac{1}{2}(s-3)}}{\sqrt{\pi}} \Gamma(s/2),$$

analytic for $\mathrm{Re}(s) > 0$
and

$$F_{21}(s) = \int_0^\infty x^{s-1} f_{21}(x)\, dx = \frac{2^{\frac{1}{2}(s-3)}}{\sqrt{\pi}} \Gamma(s/2).$$

Therefore,

$$H_1(s) = F_{11}(s)F_{21}(s) = \frac{2^{s-3}}{\pi} \Gamma^2(s/2)$$

$$h_1(y) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-s} H_1(s)\, ds$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-s} \frac{2^{s-3}}{\pi} \Gamma^2(s/2)\, ds, \qquad c > 0$$

$$= \frac{1}{2\pi} K_0(y), \qquad y > 0 \quad [6, p. 197]$$

where $K_0(y)$ is Bessel's function of the second kind with a purely imaginary argument of zero order. Similarly

$$h_2(y) = \frac{1}{2\pi} K_0(y), \qquad y < 0$$

$$h_3(y) = \frac{1}{2\pi} K_0(y), \qquad y < 0$$

$$h_4(y) = \frac{1}{2\pi} K_0(y), \qquad y > 0.$$

Therefore, $h(y) = h_1(y) + h_2(y) + h_3(y) + h_4(y)$

$$= \frac{1}{\pi} K_0(y), \qquad -\infty < y < \infty,$$

and this is the desired p.d.f. This result has been found by other methods and is given in [1, p. 1].

Example 2: *The distribution of* $\eta = \frac{\xi_1}{\xi_2}$ *where* ξ_1 *and* ξ_2 *are independent random variables with p.d.f.'s* $f_1(x)$ *and* $f_2(x)$, *respectively, where*

$$f_1(x) = f_2(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < y < \infty.$$

As in Example 1, one splits the determination of $h(y)$, the p d.f. of η, into four parts: $h_1(y)$, $h_2(y)$, $h_3(y)$, $h_4(y)$. In the notation of Example 1 it is easy to show that $H_{11}(s)$ the Mellin transform of $h_1(y)$ is

$$F_{11}(s)F_{21}(-s + 2) = \frac{2^{\frac{1}{2}(s-3)}}{\sqrt{\pi}} \Gamma(s/2) \frac{2^{\frac{1}{2}(s-3)}}{\sqrt{\pi}} \Gamma(-s/2 + 1) = \frac{1}{4} \frac{1}{\sin \frac{s\pi}{2}};$$

$$h_1(y) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-s} H(s)\, ds, \qquad 0 < c < 2,$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{1}{4} \frac{y^{-s}\, ds}{\sin \frac{s\pi}{2}}$$

$$= \frac{1}{2\pi} \frac{1}{1 + y^2}, \qquad y \geq 0.$$

Similarly

$$h_2(y) = \frac{1}{2\pi} \frac{1}{1 + y^2}, \qquad y \leq 0,$$

$$h_3(y) = \frac{1}{2\pi} \frac{1}{1 + y^2}, \qquad y \leq 0,$$

$$h_4(y) = \frac{1}{2\pi} \frac{1}{1 + y^2}, \qquad y \geq 0.$$

Therefore, $h(y) = h_1(y) + h_2(y) + h_3(y) + h_4(y)$

$$= \frac{1}{\pi} \frac{1}{1 + y^2}, \qquad -\infty < y < \infty.$$

This result has been found by other methods and given in [4, p 411].

Example 3: *F-Distribution.* Let $\xi_1, \cdots, \xi_m, \eta_1, \cdots, \eta_n$ be $(m + n)$ independ-

ent random variables, each normally distributed with mean zero and standard deviation σ. Let

$$\xi = \sum_{i=1}^{m} \xi_i^2, \qquad \eta = \sum_{j=1}^{n} \eta_j^2.$$

We want to find the p.d.f. $h(z)$ of ζ where $\zeta = \xi/\eta$. The p.d.f.'s $f(x)$ and $g(y)$ of ξ and η, respectively, are:

$$f(x) = \frac{x^{m/2-1} e^{-x/2\sigma^2}}{2^{m/2}\sigma^m \Gamma(m/2)}, \qquad x > 0,$$

and

$$g(y) = \frac{y^{n/2-1} e^{-y/2\sigma^2}}{2^{n/2}\sigma^n \Gamma(n/2)}, \qquad y > 0.$$

In this case

$$F(s) = \frac{2^{s-1}\sigma^{2s-2}\Gamma\left(s + \frac{m}{2} - 1\right)}{\Gamma(m/2)}, \qquad \text{analytic for Re } (s) > 1 - \frac{m}{2},$$

and

$$G(s) = \frac{2^{s-1}\sigma^{2s-2}\Gamma\left(s + \frac{n}{2} - 1\right)}{\Gamma(n/2)}, \qquad \text{analytic for Re } (s) > 1 - \frac{n}{2}.$$

The p.d.f. $h(z)$ has Mellin transform

$$H(s) = F(s)\, G(-s + 2)$$

$$= \frac{\Gamma\left(s + \frac{m}{2} - 1\right)\Gamma\left(-s + \frac{n}{2} + 1\right)}{\Gamma(m/2)\Gamma(n/2)}.$$

Therefore,

$$h(z) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} z^{-s} H(s)\, ds, \qquad -\frac{m}{2} + 1 < c < \frac{n}{2} + 1,$$

$$= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma(m/2)\Gamma(n/2)} \frac{z^{m/2-1}}{(z+1)^{\frac{1}{2}(m+n)}}, \qquad z > 0.$$

A convenient way of carrying out the inversion is to use formula (*d*) in Table 1.

In a similar way one can find Student's distribution, i.e., the distribution of $\zeta = \xi_0/\eta$, where $\eta = \sqrt{\sum_1^n \xi_\nu^2/n}$, and where $\xi_0, \xi_1, \cdots, \xi_n$ are $n + 1$ independent random variables each having the distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, \qquad -\infty < x < \infty.$$

It should be mentioned in conclusion that the Mellin transform is a natural tool to use in situations involving the products and quotients of independent uniformly distributed random variables, or in finding products and/or quotients and/or Beta-distribution. In such cases formulae (b), (c) and (d) in Table 1 are useful.

REFERENCES

[1] C. C. CRAIG, "On the frequency function of xy," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 1–15.

[2] H. CRAMÉR, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics, No. 36, Cambridge, 1937.

[3] H CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[4] J H CURTISS, "On the distribution of the quotient of two chance variables," *Annals of Math Stat*, Vol 12 (1941), pp 409–421

[5] E. V HUNTINGTON, "Frequency distribution of product and sum," *Annals of Math Stat*, Vol 10 (1939), pp. 195–198.

[6] E C. TITCHMARSH, *Introduction to the Theory of Fourier Integrals*, Clarendon Press, Oxford, 1937.

THE ESTIMATION OF LINEAR TRENDS

By G. W. Housner and J. F. Brennan

California Institute of Technology

1. Summary. This paper deals with the problem of bivariate regression where both variates are random variables having a finite number of means distributed along a straight line. A regression statistic is derived which is independent of change in scale so that a prior knowledge of the frequency distribution parameters is not required in order to obtain a unique estimate. The statistic is shown to be consistent. The efficiency of the estimate is discussed and its asymptotic distribution is derived for the case when the random variables are normally distributed. A numerical example is presented which compares the performance of the statistic of this paper with that of other commonly used statistics In the example it is found that the method of estimation proposed in this paper is more efficient.

2. Introduction. A problem that often arises in statistical work is the estimation of linear trends. In the general problem it is known or presumed that a linear functional relation exists among a set of variables of the form,

$$a + b_1X + b_2Y + b_3Z \cdots = 0$$

The observed values of the variables are of the form

$$x_{ik} = X_i + \epsilon_{ik}, \qquad y_{ik} = Y_i + \eta_{ik}, \quad \text{etc.}$$

That is, the x_{ik} are random variables with means X_i and $k = 1, 2, \cdot \quad N_i$ observed values of x are associated with the mean X_i. The ordering of the X_i is according to magnitude. Similarly there are the observed values y_{ik}, z_{ik} and so forth. The ϵ_{ik} are random variables, with the same distribution for all i, with zero means. On the basis of a sample $O_n(x_{ik}, y_{ik}, z_{ik}, \cdots)$ it is desired to estimate the coefficients $a, b_1, b_2, b_3, \cdots$. One method used to estimate the coefficients is that of "weighted regression" which is essentially an application of the method of least squares. The problem has been studied by R. Allen, A. Wald and others.[1] The chief difficulty has been that the proposed methods of estimation require an a priori knowledge of the variances of the random variables. Wald has proposed a statistic which avoids this difficulty but which may have a relatively low efficiency in cases often encountered in practice In this paper there is described a bivariate statistic which appears to have comparatively high precision and which does not require prior knowledge of the variances of the random variables. A numerical example is given at the end of the paper to illustrate the comparative performances of different methods of estimation.

[1] For a brief history of work done on this problem see the paper by A Wald in the *Annals of Math. Stat.*, Vol. 11 (1940), p 284.

3. The Regression statistic. In the case of the bivariate problem, consider a sample

$$O_n(x_{ik}, y_{ik}), i = 1, 2, \cdots, n$$

and

$$k = 1, 2, \cdots, N_i,$$

where N_i sample values x_i, y_i are distributed about mean X_i, Y_i. Let the means be related by $Y_i = a + bX_i$ and let the random variables x_i be independent and have the same frequency distribution with variance σ_x^2 for all i and the random variables y_i have independent frequency distributions with variance σ_y^2 the same for all i. An appropriate statistic for estimating b is obtained by noting that a pair of sample points (x_{ik}, y_{ik}), (x_{jl}, y_{jl}) gives a sample value of the change in y corresponding to a change in x. It may thus be said that a sample value of b is

$$\hat{b}_{ik,jl} = \frac{y_{ik} - y_{jl}}{x_{ik} - x_{jl}}. \tag{1}$$

Making use of the fact that

$$y_{ik} = a + bx_{ik} + \eta_{ik} - b\epsilon_{ik} \tag{2}$$

equation (1) may be written

$$(x_{ik} - x_{jl})\,\hat{b}_{ik,jl} = (x_{ik} - x_{jl})\,b + (\eta_{ik} - \eta_{jl}) - b(\epsilon_{ik} - \epsilon_{jl})$$

Summing this equation over all combinations of points there is obtained

$$b = \frac{\sum_i\sum_j\sum_k\sum_l (y_{ik} - y_{jl})}{\sum_i\sum_j\sum_k\sum_l (x_{ik} - x_{jl})} - \frac{\sum_i\sum_j\sum_k\sum_l ((\eta_{ik} - \eta_{jl}) - b(\epsilon_{ik} - \epsilon_{jl})}{\sum_i\sum_j\sum_k\sum_l (x_{ik} - x_{jl})}. \tag{3}$$

The summations in the above expression are to be carried out for

$$l = 1, 2, \cdots, N_j;\ k = 1, 2, \cdots, N_i;\ j = 1, 2, \cdots, (i-1),\ i = 1, 2, \cdots, n.$$

The first term on the right side of equation (3) is an estimate of b and the second term represents the deviation of the estimate from the true value Accordingly, we take as an estimate of b the statistic

$$\hat{b} = \frac{\sum_i\sum_j\sum_k\sum_l (y_{ik} - y_{jl})}{\sum_i\sum_j\sum_k\sum_l (x_{ik} - x_{jl})}. \tag{4}$$

This requires, of course, that the denominator be not equal to zero. Summing out the subscripts k and l reduces (4) to

$$\hat{b} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{i-1} N_i N_j(\bar{y}_i - \bar{y}_j)}{\sum_{i=1}^{n}\sum_{j=1}^{i-1} N_i N_j(\bar{x}_i - \bar{x}_j)}$$

where $\bar{y}_i$ is the mean value of the y_{ik} and so forth. Summing out the subscript j gives

$$\hat{b} = \frac{\sum_i \left(N_i \bar{y}_i \sum_1^{i-1} N_j - N_i \sum_1^{i-1} N_j \bar{y}_j\right)}{\sum_i \left(N_i \bar{x}_i \sum_1^{i-1} N_j - N_i \sum_1^{i-1} N_j \bar{x}_j\right)}. \tag{5}$$

This expression may be put in a more convenient form by using the identity

$$\sum_{i=1}^{n} \left(N_i \sum_1^{i-1} N_j \bar{y}_j\right) = \sum_{i=1}^{n} \left(N_i \bar{y}_i \sum_{i+1}^{n} N_j\right) = \sum_{i=1}^{n} \left(N_i \bar{y}_i \left(\sum_1^{n} N_j - \sum_1^{i} N_j\right)\right).$$

With this substitution equation (5) becomes

$$\hat{b} = \frac{\sum_{i=1}^{n} \left[N_i \bar{y}_i \left(\sum_1^{n} N_j - 2 \sum_1^{i} N_j + N_i\right)\right]}{\sum_{i=1}^{n} \left[N_i \bar{x}_i \left(\sum_1^{n} N_j - 2 \sum_1^{i} N_j + N_i\right)\right]}. \tag{6}$$

This is the statistic for estimating the linear trend of bivariate data. It may be noted that its derivation is not based on the notion of fitting a line to the sample points. A line $y = \hat{a} + \hat{b}x$ may be fitted to the sample points by making it pass through the mean of the sample points, that is, by using the following estimate:

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$

where $\bar{y}$ and $\bar{x}$ are the means of all the y_{ik} and x_{ik} respectively.

4. Consistency of the estimate. Having established the statistics $\hat{b}$ and $\hat{a}$ it is desirable to examine the consistency and efficiency of the estimates, particularly for $\hat{b}$. To determine that $\hat{b}$ is a consistent estimate we investigate the behavior of (6) as the number of sample points increases, that is, as the $N_i \to \infty$. We wish first to establish the following identity. Consider the sum of the following array of terms:

$$\begin{array}{l} N_1(N_1 + N_2 + \cdots + N_n) \\ N_2(N_1 + N_2 + \cdots + N_n) \\ \vdots \\ N_n(N_1 + N_2 + \cdots + N_n) \end{array}$$

The sum may be written $\sum_1^n N_i \sum_1^n N_j$. Since the array is skew symmetrical the expression $2 \sum_1^n N_i \sum_1^i N_j$ also gives the sum of the array except for the fact that the terms along the principal diagonal are counted twice We have, therefore

$$\sum_1^n N_i \sum_1^n N_j = 2 \sum_1^n N_i \sum_1^i N_j - \sum_1^n N_i^2.$$

Rearranging terms we obtain the identity

$$\sum_1^n \left[N_i\left(\sum_1^n N_j - 2\sum_1^i N_j + N_i\right)\right] = 0 \tag{7}$$

Now substituting (2) into (6) and making use of (7) there is obtained,

$$\hat{b} = b + \frac{\sum_1^n \left[N_i\left(\sum_1^n N_j - 2\sum_1^i N_j + N_i\right)(\bar{\eta}_i - b\bar{\epsilon}_i)\right]}{\sum_1^n \left[N_i\left(\sum_1^n N_j - 2\sum_1^i N_j + N_i\right)\bar{x}_i\right]} \tag{8}$$

The $\bar{\eta}_i$ and $\bar{\epsilon}_i$ are random variables with zero means so that as $N_i \to \infty$ the sample means $\bar{\eta}_i$ and $\bar{\epsilon}_i$ converge in probability to zero. As $N_i \to \infty$, $\bar{x}_i$ converges in probability to its mean X_i. In view of (7) and that the denominator in (8) is not equal to zero the last term in (8) converges in probability to zero and $\hat{b} \to b$. The estimate is therefore consistant. A similar argument also shows the estimate $\hat{a}$ to be consistent

5. Efficiency of the estimate. A general investigation of the efficiency of the estimate $\hat{b}$ is beyond the scope of this paper. We may note, however, that the efficiency of the estimate can be made to depend upon the grouping of the data, that is, the optimum efficiency of the estimate may depend upon the omission of some of the pairs $(y_{ik} - y_{jl})$ from the estimate. The maximum efficiency is obtained for $\hat{b}$ when the second term in (3) is minimized. This requires prior knowledge of the frequency distribution of the random variables x and y; however, in applications a recognition of (3) may often indicate a practical method of increasing the efficiency.

In what follows we make an investigation of the precision of the estimate $\hat{b}$ for a special case which is of some practical interest. Let x and y be random variables as defined in the first part of the paper and consider the new variables defined by $\hat{b} = \frac{v}{u}$ that is,

$$u = \sum_1^n \left[N_i\left(\sum_1^n N_j - 2\sum_1^i N_j + N_i\right)\bar{x}_i\right]$$

$$v = \sum_1^n \left[N_i\left(\sum_1^n N_j - 2\sum_1^i N_j + N_i\right)\bar{y}_i\right].$$

The random variables u and v are then independently distributed with joint probability element $f(u)\,f(v)\,du\,dv$. Making the change of variable $u = r\cos\theta$, $v = r\sin\theta$ the probability element becomes $f(r, \theta)dr\,d\theta$ where $\tan\theta = u$. Integrating out the variable r gives the probability element for θ. In what follows we investigate the distribution of θ for the case where x and y are normally distributed with the same variance. Since u and v are linear functions of x and y respectively they are also normally distributed with the same standard deviation.

We designate the means of u and v by m_1 and m_2 respectively and the standard deviation by σ The probability element in u and v is then

$$\frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[(u-m_1)^2+(v-m_2)^2]\right\} du\, dv \tag{9}$$

Changing variables to r, θ and setting $m_1 = \bar{r}\cos\bar{\theta}$, $m_2 = \bar{r}\sin\bar{\theta}$ we obtain the following probability element:

$$\frac{r}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}[(r\cos\theta-\bar{r}\cos\bar{\theta})^2+(r\sin\theta-\bar{r}\sin\bar{\theta})]\right\} dr\, d\theta$$

Completing the square in r and substituting $\phi = \theta - \bar{\theta}$ there is obtained

$$\frac{r}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(r-\bar{r}\cos\phi)^2\right\} \exp -\tfrac{1}{2}\left(\frac{\bar{r}\sin\phi}{\sigma}\right)^2\Big\} dr\, d\phi \tag{10}$$

To integrate out r make further change of variable

$$t = \frac{r}{\sigma} - \frac{\bar{r}}{\sigma}\cos\phi$$

Setting $\frac{\bar{r}}{\sigma}\cos\phi = w$ for convenience in notation there is obtained

$$\left(\frac{1}{2\pi}\, t\exp\left\{-\frac{t^2}{2}\right\} + \frac{w}{2\pi}\exp\left\{-\frac{t^2}{2}\right\}\right)\exp\left\{-\tfrac{1}{2}\left(\frac{\bar{r}^2}{\sigma^2}-w^2\right)\right\} dt\, d\phi$$

The variable t is to be integrated out of this expression. The corresponding limits of integration are exhibited by

$$\begin{aligned}\frac{w}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\left(\frac{\bar{r}^2}{\sigma^2}-w^2\right)\right\}&\left(\frac{1}{\sqrt{2\pi}}\int_{-w}^{+\infty}\frac{t}{w}\exp\left\{-\frac{t^2}{2}\right\}dt\right.\\ &\left.+\frac{1}{\sqrt{2\pi}}\int_{-w}^{+\infty}\exp\left\{-\frac{t^2}{2}\right\}dt\right)d\phi.\end{aligned} \tag{12}$$

Now as the number of points in the original sample increases the value of $\bar{r}$ also increases and as $\frac{\sigma}{\bar{r}} \to 0$, with $|\phi| < \frac{\pi}{2}$, the value of $w \to \infty$. In this case then (12) approaches asymptotically to

$$\frac{\bar{r}}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\sin\phi^2}{\sigma/\bar{r}}\right)^2\right\} d(\sin\phi).$$

As $\sigma/\bar{r} \to 0$ this distribution shows that ϕ converges in probability to zero and that the distribution approaches asymptotically to the normal form

$$\frac{1}{\sqrt{2\pi}\sigma/\bar{r}}\exp\left\{-\tfrac{1}{2}\left(\frac{\phi}{\sigma/\bar{r}}\right)^2\right\} d\phi. \tag{13}$$

It is required then to examine the conditions under which $\sigma/\bar{r}$ assumes small values. If the variance of the original variables x_i and y_i is designated by σ_1^2

then since u and v are linear functions of x_i and y_i respectively the variance of u and of b is

$$(14) \qquad \sigma^2 = \sigma_1^2 \sum_1^n \left\{ \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \right]^2 \left(\frac{1}{N_i} \right) \right\}.$$

Now $\bar{r}^2$ is the sum of the squares of the means of u and v so that

$$(15) \qquad \bar{r}^2 = (1 + b^2) \left\{ \sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) X_i \right] \right\}^2.$$

Dividing (14) by (15) we obtain

$$(16) \qquad \left(\frac{\sigma}{\bar{r}} \right)^2 = \frac{\sigma_1^2}{1 + b^2} \frac{\sum_1^n \left\{ \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) \right]^2 \left(\frac{1}{N_i} \right) \right\}}{\left\{ \sum_1^n \left[N_i \left(\sum_1^n N_j - 2 \sum_1^i N_j + N_i \right) X_i \right] \right\}^2}.$$

Inspection of (16) indicates that as the number of sample points N_i increases the value of $\left(\frac{\sigma}{\bar{r}} \right)^2$ decreases rapidly. To illustrate this we examine some particular cases Consider first the case of four equally spaced means $X_i = 3i\sigma_1$, $(i = 1, 2, 3, 4)$ and let there be one sample point for each mean $(N_i = 1)$. With these values there is obtained,

$$\left(\frac{\sigma}{\bar{r}} \right)^2 = \frac{0.022}{1 + b^2}.$$

For $b = 1$ the range $-9° < \phi < +9°$ includes 95% of the population defined by (13). As the number of points N_i is increased or as the number of means X_i is increased the value of $\left(\frac{\sigma}{\bar{r}} \right)^2$ decreases rapidly. Consider now eight equally spaced means $X_i = 3_i\sigma_1$, $(i = 1, 2, \cdots, 8)$ with again one sample point for each mean $(N_i = 1)$. With these values there is obtained

$$\left(\frac{\sigma}{\bar{r}} \right)^2 = \frac{0.00045}{1 + b^2}.$$

For $b = 1$ the range $-1° < \phi < +1°$ includes 95% of the population defined by (13).

It is clear that a very high degree of precision is obtained with the estimate $\hat{b}$ when there is a considerable number of sample points. However, this will also be true in general of other statistics and it is really of interest to compare precisions in those cases where the statistics have a relatively low precision A detailed comparison is beyond the scope of this paper. However, a direct comparison can be made very easily in the particular case when x_i is a fixed variate

and only y_i is a random variable. For the sake of simplicity, let each $N_i = 1$ then the statistic for estimating b is

$$\hat{b} = \frac{\sum_1^n i(y_i - \bar{y})}{\sum_1^n i(x_i - \bar{x})} = \frac{\sum_1^n y_i(i - \bar{i})}{\sum_1^n x_i(i - \bar{i})}. \tag{17}$$

Since $\hat{b}$ is a linear function of the y_i by a well known theorem its variance is

$$\sigma_{\hat{b}}^2 = \sigma_y^2 \sum_1^n \left(\frac{i - \bar{i}}{\sum_1^n x_i(i - \bar{i})} \right)^2. \tag{18}$$

The customary least squares regression line of y on x gives for the estimate of b and its variance

$$\hat{b}_R = \frac{\sum_1^n y_i(x_i - \bar{x})}{\sum_1^n x_i(x_i - \bar{x})} \qquad \sigma_{\hat{b}_R}^2 = \sigma_y^2 \sum_1^n \left(\frac{x_i - \bar{x}}{\sum_1^n x_i(x_i - \bar{x})} \right)^2.$$

In the particular case when the x_i are equally spaced, $x_i = ci + d$, the estimates $\hat{b}$ and $\hat{b}_R$ are identical:

$$\hat{b} = \hat{b}_R = \frac{12}{cn(n^2 - 1)} \sum_1^n y_i(i - \bar{i}). \tag{19}$$

6. Numerical example. From a practical point of view the case where x and y are random variables is of greater interest than where x is a fixed variate. We give a numerical example of this case comparing the statistic $\hat{b}$ with several other statistics. Consider the case where there is one sample point for each mean X_i. We shall evaluate the following:

1) The statistic of this paper which for this case is

$$\hat{b}_1 = \frac{\sum_1^n y_i(i - \bar{i})}{\sum_1^n x_i(i - \bar{i})}.$$

2). The statistic obtained by minimizing the sum of the squares of the y deviations only

$$\hat{b}_2 = \frac{\sum_1^n y_i(x_i - \bar{x})}{\sum_1^n x_i(x_i - \bar{x})}.$$

3). The statistic obtained by minimizing the sum of the squares of the orthogonal deviations

$$\hat{b}_3 = \frac{\sum_1^n (y_i - \bar{y})^2 - \sum_1^n (x_i - \bar{x})^2 + \left[n \sum_1^n (y_i - \bar{y})^2 - n \sum_1^n (x_i - \bar{x})^2 + 4\left(\sum_1^n (y_i - \bar{y})(x_i - \bar{x})\right)^2\right]^{\frac{1}{2}}}{\sum_1^n (y - \bar{y})(x - \bar{x})}$$

TABLE I

Set	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
1	1.1	1.4	2 4	2 0	3 0	2.7	3.6	4.3
2	1.2	1.4	2 2	2.0	3 4	3.1	3 8	4.2
3	1.0	1.4	1.6	2.1	2.8	3 2	4.4	4.3
4	0.6	0 7	1.8	2.0	3 3	2 6	3.8	4 0
5	0.7	1 4	1.7	1 7	2 7	3.4	4 1	4.1
6	1.0	1.2	1 6	2.1	2.9	2.6	3 6	4 0
7	1 3	0 7	1 7	2 1	2.7	2.9	4.0	3.6

TABLE II

Set	b_1	b_2	b_3	b_4
1	1.160	1.068	1 220	1 162
2	1 056	1.009	1.059	1.027
3	0 860	0.843	0.803	0.870
4	0 946	0.896	0 924	0.830
5	0.875	0 867	0 913	1.000
6	0 978	0.939	0.981	0.846
7	1 044	0.959	1 045	1.000
Mean . . .	0 990	0.940	0 996	0 962
7 × Sample Var .	0 0686	0 0373	0 1058	0.0834

4). The statistic proposed by Wald[2]

$$\hat{b}_4 = \frac{\sum_1^{n/2} y_i - \sum_{n/2}^n y_i}{\sum_1^{n/2} x_i - \sum_{n/2}^n x_i}.$$

We apply these statistics to sample data having four means $X_i = i$ and $Y_i = i$, $(i = 1, 2, 3, 4)$. By means of a table of random numbers seven sets of data were

[2] Loc. cit.

obtained, each set having one sample point corresponding to each mean. These sample points are described by Table I where it will be noted that the sample points were drawn from a discrete distribution. The estimates obtained from the four statistics are exhibited in Table II.

If the 28 sample points are treated as a single set of data and the four statistics in their appropriate forms are applied, there is obtained the following set of estimates:

$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$
0.9768	0.9183	0.9786	0.9496

The preceding computations show that the estimate $\hat{b}_2$ is inferior to the other estimates, as would be expected. The estimate $\hat{b}_3$ is most accurate when the 28 sample points are treated as a single set of data with the estimate $\hat{b}_1$ being only very slightly less accurate, $\hat{b}_1 = 0.9768$ as compared to $\hat{b}_3 = 0.9786$. When the individual sets of sample points 1 to 7 are considered it is seen that the estimate $\hat{b}_1$ is most accurate with the estimate $\hat{b}_3$ rather less accurate, the estimate $\hat{b}_1$ is more precise than $\hat{b}_3$, the sample variances being in the ratio $0.0686 \div 0.1058 = 0.65$. From a practical viewpoint we may also point out that the computation of $\hat{b}_1$ requires very much less labor than the computation of $\hat{b}_3$.

ON THE EFFECT OF DECIMAL CORRECTIONS ON ERRORS OF OBSERVATION

By Philip Hartman and Aurel Wintner

The Johns Hopkins University

1. Summary. Let t be the true value of what is being measured and suppose that the error of observation is a symmetric normal distribution of standard deviation σ. The "rounding-off" error due to the reading of measurements to the nearest unit has a distribution and an expected value depending on t and σ. It is shown that, for a fixed $\sigma > 0$, the expected value of the decimal correction, $r(t; \sigma)$, is an analytic function of t which is odd, of period 1, positive for $0 < t < \frac{1}{2}$, and has a convex arch as its graph on $0 \leqq t \leqq \frac{1}{2}$. Furthermore, if $0 < t < \frac{1}{2}$, both $r(t; \sigma)$ and its maximum value, $\operatorname*{Max}_t r(t, \sigma)$, are decreasing functions of σ.

2. Introduction. Let X be an error of observation and let $\phi(x)$ denote the density of probability of the distribution of X. In particular,

$$\int_{-\infty}^{+\infty} \phi(x)\, dx = 1, \qquad \text{where } \phi(x) \geqq 0 \tag{1}$$

If t is any fixed number, the density of probability of the distribution of $X + t$ is $\phi(x - t)$.

Besides the "instrumental error of observation", X, there is another error, that of the "rounding-off", which is carried along in the registration of the measurements. It is introduced by the circumstance that, if $\cdots, b, a$ are digits, and if b denotes the last digit considered, then decimal fractions such as $\cdots ba$ and $\cdots ba \cdots$ are registered as $\cdots b$ if $a < 5$ and as $\cdots (b + 1)$ if $a > 5$. Let the unit, in which the measurements are expressed, be so chosen that the first digit neglected becomes the first digit following the decimal point, i.e., that the error of the "rounding-off" is between $\pm\frac{1}{2}$. Then, if t denotes the *true* value of what is being measured, the remark made after (1) shows that *the probability that the error of the decimal corrections be less than* x is given by

$$\sum_{n=-\infty}^{\infty} \int_{n-\frac{1}{2}}^{n-\frac{1}{2}+x} \phi(u - t)\, du,$$

if $|x| \leqq \frac{1}{2}$, whereas this probability is 0 or 1 according as $x < -\frac{1}{2}$ or $x > \frac{1}{2}$. Since the last series can be written in the form

$$\sum_{n=-\infty}^{\infty} \int_{-\frac{1}{2}}^{-\frac{1}{2}+x} \phi(u + n - t)\, du = \int_{-\frac{1}{2}}^{-\frac{1}{2}+x} \sum_{n=-\infty}^{\infty} \phi(u + n - t)\, du, \quad (\phi \geqq 0), \tag{2}$$

it follows that the density of probability of the error due to the decimal corrections is

$$\sum_{n=-\infty}^{\infty} \phi(x + n - t) \text{ if } |x| < \tfrac{1}{2}, \text{ and } 0 \text{ if } |x| > \tfrac{1}{2} \tag{3}$$

Consequently, *if $r = r(t)$ denotes the expected value of the decimal error induced on the "true" value, t, of the observations, then*

$$(4) \qquad r(t) = \int_{|x|<\frac{1}{2}} x \sum_{n=-\infty}^{\infty} \phi(x + n - t)\, dx.$$

Formula (4) is known[1] It is usually based on its intuitive interpretation which results if, on the one hand, (4) is written in the form

$$(5) \qquad r(t) = \int_{-\infty}^{\infty} s(x)\phi(x - t)\, dx,$$

where

$$(6) \qquad s(x) = x \text{ if } -\tfrac{1}{2} < x < \tfrac{1}{2} \text{ and } s(x) = s(x+1), \qquad -\infty < x < \infty,$$

and, on the other hand, the periodic function (6) is thought of as representing the *uniform* distribution of the error of "rounding-off" over the *arithmetical continuum* over a period,

$$|x - n| < \tfrac{1}{2}, \qquad (n = 0, \pm 1, \cdots),$$

on the x-axis. Needless to say, the specification of $s(x)$ at the points $x = n + \frac{1}{2}$, which are disregarded in the definition (6), is immaterial, since $s(x)$ occurs in (5) only as an integrable weight-factor, isolated values of which do not influence the integral.

It follows at once from (1), (5) and the continuity (almost everywhere) of (6), that $r(t)$ is continuous.

3. Fourier analysis of $r(t)$. Since the Fourier expansion of the periodic function (6) is

$$(7) \qquad s(x) = -\pi^{-1} \sum_{n=1}^{\infty} (-1)^n n^{-1} \sin 2\pi n x = s(x \pm 1) = \cdots, \qquad (|x| < \tfrac{1}{2}),$$

it follows from (5) that[2]

$$(8) \qquad r(t) = -\pi^{-1} \sum_{n=1}^{\infty} (-1)^n n^{-1} \int_{-\infty}^{\infty} \phi(x) \sin 2\pi n(x + t)\, dx.$$

Hence, if the sine in (8) is expressed in terms of $2\pi nx$ and $2\pi nt$,

$$(9) \qquad \pi r(t) = -\sum_{n=1}^{\infty} (-1)^n n^{-1}(a_n \cos 2\pi nt + b_n \sin 2\pi nt),$$

---

[1] F. Zernike, "Wahrscheinlichkeitsrechnung und mathematische Statistik," *Handbuch der Physik*, Vol 3 (1928), pp 475–476.

[2] In view of (1), the term-by-term integration leading from (5) to (8) is justified by the fact that the partial sums of the series (7) are uniformly bounded Correspondingly, the above deduction of (9) and (10) from (4) is equivalent to an application of Poisson's summation formula. In this regard, cf. A Wintner, "The sum formulae of Euler-Maclaurin and the inversions of Fourier and Mobius," *Am. Jour of Math*, Vol. 69 (1947), pp. 685–708, the end of §1 (p 687) and its application on p. 697.

where

(10) $$b_n + ia_n = \int_{-\infty}^{\infty} \phi(x) \exp(2\pi i n x)\, dx, \qquad (n = 1, 2, \cdots).$$

Let it be assumed that positive and negative errors of observation, when of the same magnitude, are equally probable, i.e., that $\phi(x) = \phi(-x)$. Then (10) shows that a_n becomes 0. Hence, (9) reduces to

(11) $$r(t) = -\sum_{n=1}^{\infty} (-1)^n (c_n/n) \sin 2\pi n t,$$

where

(12) $$c_n = \pi^{-1} \int_{-\infty}^{\infty} \phi(x) \cos 2\pi n x\, dx = 2\pi^{-1} \int_0^{\infty}.$$

Clearly, $r(t)$ is an odd function whenever the density $\phi(x)$ is even.

4. The normal case. Suppose that $\phi(x)$ is the density of a symmetric *normal* (Gaussian) distribution. Then, if σ is the positive constant representing the standard deviation of the errors of observation,

(13) $$\phi(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}x^2/\sigma^2) \qquad (0 < \sigma < \infty).$$

It is clear from (5) and (6) that

(14) $$r(t) \to s(t) \text{ if } \sigma \to 0 \text{ in (13).}$$

Actually, all that (14) says is a triviality, according to which the total error becomes the decimal error when the measurements become infinitely sharp. In this limiting case, that is, if $r(t) = s(t)$, it is seen from (6) that the graph of the periodic function $r = r(t)$ is piecewise linear, and therefore discontinuous.

If $\sigma = 0$ is replaced by $0 < \sigma < \infty$, the jumps of $r(t)$ at $t = n - \frac{1}{2}$ disappear (cf. the end of §3) and, as will be proved below,

(I) *$r(t)$ is an analytic function which is odd, of period* 1, *and positive for* $0 < t < \frac{1}{2}$ (*hence negative for* $-\frac{1}{2} < t < 0$), *and*

(II) *the graph of* $r = r(t)$ *over the fundamental interval* $0 \leqq t \leqq \frac{1}{2}$ *is a convex arch, no matter what the value of* σ *in* (13) *may be.*

Since r now depends both on the "true" value, t, of the observations and the "precision", σ, of the measurements, let r be denoted by $r(t, \sigma)$. It will be shown that

(i) Max $r(t; \sigma)$, *where the* Max *refers to t while σ is fixed, is a decreasing function of* σ, where σ varies on the half-line $0 < \sigma < \infty$, and that, on the same half-line,

(ii) *$r(t; \sigma)$ is a decreasing function of σ at every fixed t contained in the fundamental region* $0 < t < \frac{1}{2}$.

All of this seems to be clear for physical reasons. Actually, it is easy to give examples of distribution laws, distinct from (13) for which the above assertions become false.

5. The ϑ_3-function. As is well-known,

$$\int_{-\infty}^{\infty} \exp\left(-\tfrac{1}{2}x^2/\sigma^2\right) \cos ux \, dx = (2\pi\sigma^2)^{\frac{1}{2}} \exp\left(-\tfrac{1}{2}\sigma^2 u^2\right)$$

Hence, the value of the integral (12) is q^{n^2}, if q is an abbreviation for

$$q = \exp(-2\pi^2\sigma^2) \tag{15}$$

Consequently, if $r(t, q)$ is defined, in terms of the above $r(t; \sigma)$, by placing

$$r(t, q) = r(t, \sigma) \text{ in virtue of (15)}, \tag{16}$$

then (11) shows that[3]

$$r(t, q) = -\pi^{-1} \sum_{n=1}^{\infty} (-1)^n n^{-1} q^{n^2} \sin 2\pi nt \tag{17}$$

It will be noted that the range, $0 < \sigma < \infty$, of the standard deviation is mapped by (15) on the range

$$0 < q < 1, \tag{18}$$

and that σ decreases or increases according as q increases or decreases.

Let partial differentiations with respect to t and q be denoted by primes and subscripts, respectively.

$$f' = \partial f/\partial t, \qquad f_q = \partial f/\partial q. \tag{19}$$

Thus, from (17),

$$r'(t, q) = -2 \sum_{n=1}^{\infty} (-1)^n q^{n^2} \cos 2\pi nt \tag{20}$$

and, as easily verified from (17),

$$r_q(t, q) = (-4\pi q)^{-1} r''(t, q). \tag{21}$$

Let $\theta(t, q)$ be defined by

$$\theta(t, q) = 1 + 2 \sum_{n=1}^{\infty} q^{n^2} \cos nt \tag{22}$$

(so that $\theta(t, q)$ is, in the main, the elliptic theta-function usually denoted by ϑ_3). It is known that

$$\theta(t, q) > 0 \tag{23}$$

and that[4]

$$\theta'(t, q) < 0 \text{ if } 0 < t < \pi \qquad (\text{hence, } \theta'(t, q) > 0 \text{ if } -\pi < t < 0). \tag{24}$$

The above assertions will be deduced from these facts.

---

[3] Cf. F Zernike, loc. cit

[4] For a simple proof, cf. A. Wintner, "On the shape of the angular case of Cauchy's distribution," *Annals of Math. Stat.*, Vol. 18 (1948), pp. 589–593, §6

6. **Proof of (I)–(II) and (i)–(ii).** First, it is seen from (17) and (22) that

$$r'(t, q) = 1 - \theta(2\pi t - \pi, q). \tag{25}$$

Hence,

$$r''(t, q) = -2\pi\theta'(2\pi t - \pi, q). \tag{26}$$

If (26) is compared with (24), it is seen that

$$r''(t, q) < 0 \text{ if } 0 < t < \tfrac{1}{2} \qquad (\text{hence, } r''(t, q) > 0 \text{ if } -\tfrac{1}{2} < t < 0). \tag{27}$$

Consequently, (I) and (II) follow, since, in view of (17),

$$r(\pm\tfrac{1}{2}, q) = 0 = r(0, q). \tag{28}$$

Next, (21) and (27) imply that

$$r_q(t, q) > 0 \text{ for } 0 < t < \tfrac{1}{2}. \tag{29}$$

Hence, (ii) follows from the fact that q is a decreasing function of σ.

As to (i), let $t = t(q)$ denote that (unique) t-value on $0 < t < \frac{1}{2}$ at which $r(t, q)$ assumes its maximum value, say r^q, so that

$$r^q = r(t(q), q), \qquad (0 < t(q) < \tfrac{1}{2}). \tag{30}$$

Clearly, $t = t(q)$ is the only t-value on $0 < t < \frac{1}{2}$ for which

$$r'(t, q) = 0 \tag{31}$$

Since $r'(t, q)$ possesses continuous partial derivatives with respect to t and q, and since (27) implies that its partial derivative with respect to t, namely, $r''(t, q)$, does not vanish at $t = t(q)$, it follows that the solution $t = t(q)$ of the equation (31) possesses a continuous derivative Hence, the function (30) possesses a continuous derivative with respect to q, namely,

$$\frac{dr^q}{dq} = r'(t(q), q)\,\frac{dt(q)}{dq} + r_q(t(q), q) \tag{32}$$

But since $t = t(q)$ is a solution of (31), the identity (32) can be reduced to

$$\frac{dr^q}{dq} = r_q(t(q), q), \qquad (0 < t(q) < \tfrac{1}{2}).$$

Consequently, (i) follows from (29), since q is a decreasing function of σ.

WEIGHING DESIGNS AND BALANCED INCOMPLETE BLOCKS

By K. S. Banerjee

Pusa, Bihar, India

1. Introduction. Following a paper by Hotelling [1] on the weighing problem, Kishen [4] and Mood [2] furnished generalized solutions. This note consists of some additional remarks on the weighing problem when the weighing is restricted to be made on one pan.

Hotelling remarked that when the problem was to determine a particular difference or any other linear function of the weights, a different design should be sought to minimize the variance. An account of efficient designs of this kind has also been furnished in this note. The notations used by Hotelling and Mood have been used here

2. Chemical balance problem. It has been shown by Mood that when $N \equiv 0 \pmod 4$, an optimum design exists if a Hadamard matrix H_N exists, and is obtained by using any p columns of H_N. When $N \equiv i \pmod 4$, ($i = 1, 2, 3$), very efficient designs are obtained either by adding to or deleting from the rows of H_{4K}, making the resultant number of rows equal to N.

It has further been shown by Mood in connection with this class of designs that arrangements[1] are available which are more efficient than the one obtained by repeating the row of ones. As a matter of fact, if any row other than the row of ones be repeated, this will lead to a design of the same efficiency as in the case of repeated addition of the row of ones, for, the determinant of $X'X$ will remain exactly identical That this is so, will be clear from the following properties showing the connection of the matrix X with the determinant $|a_{ij}|$:

(i) Any two rows of the matrix X can be interchanged without changing the determinant $|a_{ij}|$.

(ii) Any two columns of the matrix X can be interchanged without changing the determinant $|a_{ij}|$

(iii) The signs of all the elements in a column of the matrix X may be changed without changing the determinant $|a_{ij}|$

3. Spring balance problem. Mood has exhaustively discussed the designs when $N > p$ Efficient designs under this class will, however, be available from the arrangements afforded by balanced incomplete block designs discussed in [3]. These designs will be represented by certain of the efficient submatrices of the P_k of Mood.

Usually v and b are used to denote respectively the number of varieties and the number of blocks in the above mentioned designs. Here v will take the place of

[1] This had been independently shown by me before the paper of A. M. Mood was brought to my notice by H Hotelling

p, the number of objects to be weighed and b that of N, the number of weighings that can be made The matrix $X'X$ in this case will take the form

(1)
$$\begin{bmatrix} r & \lambda & \lambda & & \lambda \\ \lambda & r & \lambda & \cdot & \lambda \\ \lambda & \lambda & r & \cdot & \lambda \\ & \cdot & & \cdot\,\cdot & \cdot \\ & \cdot & & \cdot & \\ \lambda & \lambda & \lambda & \cdot & r \end{bmatrix}$$

The variance of the estimated weight of each of the p objects for such a design can be easily seen to be

(2)
$$\frac{r + \lambda(p - 2)}{(r - \lambda)\{r + \lambda(p - 1)\}}\,\sigma^2 \qquad \text{for zero bias,}$$

where p is the number of objects to be weighed and r and λ have meanings similar to those in connection with balanced incomplete block designs; that is, r is the number of times each object is weighed, and λ is the number of times each pair of objects is weighed together

Though the *minimum minimorum* of σ^2/N can never be attained by the objects to be weighed under such designs, σ^2/N may however be kept as the standard with which the efficiency of a given design may be calculated. The efficiency of the above design will therefore for zero bias be

(3)
$$\frac{(r - \lambda)\{r + \lambda(p - 1)\}}{N\{r + \lambda(p - 2)\}}.$$

The identities well known in the theory of balanced incomplete blocks,

$$bk = vr, \qquad \lambda(v - 1) = r(k - 1),$$

may, upon replacing b by N and v by p to accord with the notation of weighing designs, be written

$$r = Nk/p, \qquad \lambda = r(k - 1)/(p - 1)$$

Upon substituting these in (3) we obtain the efficiency factor in the form

(4)
$$\frac{k^2(p - k)}{p(pk - 2k + 1)},$$

where k is the number of plots per block or the number of objects that can be weighed at a time.

If instead of adopting repetitions of P_K, only $\binom{p}{K}$ weighings be made in all, the efficiency factor calculated for such a combinatorial design would be

$$\frac{(r - \lambda)\{r + \lambda(v - 1)\}}{b\,\{r + \lambda(v - 2)\}}, \qquad \text{for zero bias}$$

where

$$r = \binom{v-1}{K-1}, \qquad \lambda = \binom{v-2}{K-2}$$

and $b = \binom{v}{K}$. The above expression on simplification reduces to (4).

It will be noticed that the efficiency of such designs depends only upon the total number of objects to be weighed and the number of such objects that can be weighed at a time.

These designs have the advantage that all the weights are estimated with equal precision. If a slightly larger number of weighing than what is afforded by the number of blocks in a balanced incomplete block design has to be made, all the objects may be weighed together and this weighing be repeated as many times as required. This will be equivalent to the repeated addition of the row of ones. The repetition of the row of ones in particular is necessary to make the weights estimable with equal precision, which however, may be demanded at times as a matter of necessity in certain experiments. Otherwise, any other single row or different rows of the matrix X may be repeated, making the number of rows of the matrix X equal to the number of weighings proposed to be made in all.

From the practical point of view also, it will be advantageous to connect the designs for weighing with the already existing balanced incomplete block designs, which have been highly developed in recent years and are being extensively used in agro-biological investigations.

4. Spring balance design for small p. Under this class of designs, Mood has found the most efficient design for $p = 7$. It is given by

$$L_7 = \begin{bmatrix} 1&0&1&0&1&0&1 \\ 0&1&1&0&0&1&1 \\ 0&0&0&1&1&1&1 \\ 1&1&0&0&1&1&0 \\ 0&1&1&1&1&0&0 \\ 1&0&1&1&0&1&0 \\ 1&1&0&1&0&0&1 \end{bmatrix}$$

This L_7 is easily recognized to be the design for $k = 4$, $b = 7$, $v = 7$, $r = 4$, $\lambda = 2$, given by an orthogonal series [3]. It is therefore seen that Hadamard matrices will lead to a new method of constructing balanced incomplete block designs of a certain class. For example H_{16} and H_{20} will lead respectively to the designs for $k = 8$, $b = 15$, $v = 15$, $r = 8$, $\lambda = 4$ (or for $k = 7$, $b = 15$, $v = 15$, $r = 7$, $\lambda = 3$) and for $k = 10$, $b = 19$, $v = 19$, $r = 10$, $\lambda = 5$ (or $k = 9$, $b = 19$, $v = 19$, $r = 9$, $\lambda = 4$). These designs also satisfy the condition of maximum

efficiency, by virtue of the fact that $|L_N|$ will have the value

$$(N+1)^{\frac{1}{2}(N+1)}/2^N,$$

as shown by Mood.

6. Determination of a linear function of the objects. An orthogonalized design which is cent percent efficient to determine individually the weight of p unknown objects is not necessarily the design of maximum efficiency for the estimation of a linear function of the objects. To illustrate this, let there be three objects, the weights O_1, O_2, O_3, of which have to be estimated on a balance corrected for zero bias and let us, for this purpose, concentrate on the design characterized by the matrix given below.

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}. \tag{5}$$

As has been indicated in the previous papers, the variance of each of the unknown objects comes out to be $\frac{1}{4}\sigma^2$, which is the *minimum minimorum* and as such the above design enjoys the cent percent efficiency, when the question of individual estimation is concerned But in estimating a linear function of the objects, for instance the total weight, designs more efficient than this are available

The variance of $l_1O_1 + l_2O_2 + l_3O_3$ is known to be

$$\sum_{i,j=1}^{3} l_i l_j C_{ij} \sigma^2 \tag{6}$$

where C_{ij} denotes the elements of the matrix reciprocal to the matrix $X'X$ As the above design furnishes the estimates of the unknown objects orthogonally, the variance of the estimated total weight of the three objects will be given by $\frac{3}{4}\sigma^2$ If, however, the design given by the matrix

$$x = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{7}$$

be adopted, the variance of the estimate of the total weight may be easily seen to be $(3/7)\sigma^2$, by putting $l_1 = l_2 = l_3 = 1$. $(3/7)\sigma^2$ is evidently less than $\frac{3}{4}\sigma^2$. Therefore with four weighings, the design characterized by (7) is more efficient in estimating the total weight than that characterized by (5) A still more efficient design for getting the total weight is simply to weigh all the objects together four times

6. Designs with arrangements afforded by balanced incomplete blocks. The necessity for an efficient design to estimate any linear function of the objects

(or to be precise, say to estimate the total weight) will perhaps arise only when the objects cannot all be weighed at a time collectively on a single pan. Here also, an efficient design under the supposition that all the objects cannot be weighed together is afforded by the arrangements in balanced incomplete blocks. In such a design, the diagonal elements in the matrix reciprocal to $X'X$ will be all positive and equal to

$$\frac{r + \lambda(p-2)}{(r-\lambda)\{r + \lambda(p-1)\}}, \tag{8}$$

while the remaining elements in the reciprocal matrix will be negative and equal to

$$\frac{-\lambda}{(r-\lambda)\{r + \lambda(p-1)\}}. \tag{9}$$

Using the generalized form of (6) and admitting of the possibility that any of the arbitrary constants l_i may be negative, the variance of the linear function $\sum_{i=1}^{p} l_i O_i$ may be easily seen to be

$$\left\{\frac{\Sigma l_i^2}{r-\lambda} - \frac{\lambda(\Sigma l_i)^2}{(r-\lambda)\{r+(p-1)\lambda\}}\right\}\sigma^2. \tag{10}$$

If, however, in the above expression, the coefficients l_i are equal to 1, (10) is the variance of the estimated total weight, and reduces to

$$\frac{p}{r + (p-1)\lambda}\sigma^2. \tag{11}$$

When there are N weighings in all, the minimum variance that can be reached is σ^2/N and will be attained, it appears, only when all the objects are weighed together and the weighing is repeated N times The efficiency of a given design may therefore be calculated with reference to σ^2/N. Remembering that the number of weighings takes the place of the number of blocks and p the place of v, the efficiency of the design will reduce to $(k/p)^2$, where k is the number of plots per block i.e the number of objects that can be weighed at a time

If, however, the combinatorial arrangement is adopted weighing all possible combinations of k objects and making $\binom{p}{k}$ weighings in all, the same efficiency as above will be obtained for such a design.

Given k, the above expression of efficiency will therefore be the deciding factor for choice between an arrangement of balanced incomplete block design and all possible combinations of k objects.

7. Design of maximum efficiency. Designs leading to the matrix $X'X$ of the type (1) have certain advantages inasmuch as the variances of the individual objects are equal, as are also the covariances between all possible pairs. The

variance of the estimated total weight in such a design is given by (11). To minimize the variance thus obtained, the expression

$$r + (p - 1)\lambda \tag{12}$$

has to be the maximum for a given value of p. In an arrangement of the balanced incomplete block type or in an arrangement with all possible combinations of k objects being weighed at a time, (12) would reduce to rk and would therefore increase with the increasing value of rk. This shows that the estimation of the total weight will have increased precision if more of the objects are weighed at a time.

If all the objects could be weighed at a time and both the pans be used for the purpose, some of the elements in the matrix X will be -1 instead of 0. This would increase the value of r but would decrease the value of λ. To devise the best possible design therefore, account will have to be taken simultaneously of r and λ.

REFERENCES

[1] HAROLD HOTELLING, "Some improvements in weighing and other experimental techniques", *Annals of Math. Stat.*, Vol. 15 (1944), pp. 297–306.

[2] A. M. MOOD, "On Hotelling's weighing problem", *Annals of Math. Stat.*, Vol. 17 (1946), pp. 432–446.

[3] R. A. FISHER AND F. YATES, *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, London, 1938, pp. 10–13.

[4] K. KISHEN, "On the design of experiments for weighing and making other types of measurements", *Annals of Math. Stat.*, Vol. 16 (1945), pp. 294–300.

[5] C. R. RAO, "On the most efficient designs in weighing", *Sankhyā*, Vol. 7 (1946), pp. 440.

BOUNDS FOR SOME FUNCTIONS USED IN SEQUENTIALLY TESTING THE MEAN OF A POISSON DISTRIBUTION[1]

By Leon H. Herbach

Brooklyn College

1. Introduction. Let $z = \log \dfrac{f(x, \lambda_1)}{f(x, \lambda_0)}$, where $f(x, \lambda_i) = (e^{-\lambda_i} \lambda_i^x)/x!$, $(i = 0, 1)$, is the elementary probability law of a Poisson variate X, under the hypothesis that the mean is equal to λ_i. Without loss of generality we shall assume $\lambda_1 > \lambda_0$.

Let H_0 be the hypothesis that the distribution of X is given by $f(x, \lambda_0)$. Wald [1, pp. 286–287] has devised general upper and lower bounds for the probability of accepting H_0, when λ is the true value of the parameter, and the sequential probability ratio test is used. This probability is called the operating-characteristic function and is designated by $L(\lambda)$. Using these results he has computed the bounds for the binomial and normal distributions [2, pp. 137–142]. We shall do the same thing for the Poisson distribution, since the restrictions [1, p 284, conditions I to III] under which these general limits are valid can rather easily be shown to apply to the Poisson distribution, if we make the further restriction that $E(z) \neq 0$.

These general results are

$$\frac{1 - B^h}{\delta A^h - B^h} \leq 1 - L(\lambda) \leq \frac{1 - \eta B^h}{A^h - \eta B^h}, \qquad \text{if } h > 0,$$

and

$$\frac{1 - A^h}{\delta B^h - A^h} \leq L(\lambda) \leq \frac{1 - \eta A^h}{B^h - \eta A^h}, \qquad \text{if } h < 0, \tag{1}$$

where α, β are probabilities of committing errors of the first and second kind respectively and

$$A = (1 - \beta)/\alpha, \qquad B = \beta/(1 - \alpha)$$

$$\eta = \operatorname*{glb}_{\zeta} \zeta E\left(e^{hz} \mid e^{hz} < \frac{1}{\zeta}\right), \qquad \zeta > 1;$$

$$\delta = \operatorname*{lub}_{\rho} \rho E\left(e^{hz} \mid e^{hz} \geq \frac{1}{\rho}\right), \qquad 0 < \rho < 1; \tag{2}$$

and h is the non-zero root of the expression, $Ee^{zt} = 1$. Hence the only remaining unknowns are η and δ

[1] The author is indebted to Professor A Wald for suggesting the problem which led to this note and for helpful discussions.

The following bounds to En, the expected number of observations required by the sequential probability ratio test defined by α, β have been derived [1, pp. 143–147]:

$$\frac{L(\lambda)(\log B + \xi') + [1 - L(\lambda)] \log A}{Ez} \underset{\geq}{\leq} En \underset{\geq}{\leq} \frac{L(\lambda) \log B + [1 - L(\lambda)](\log A + \xi)}{Ez},$$

the upper or lower inequality signs holding according as $Ez > 0$ or $Ez < 0$, where

$$\xi' = \operatorname*{Min}_r E(z + r \mid z + r \leq 0), \tag{3}$$

and

$$\xi = \operatorname*{Max}_r E(z - r \mid z - r \geq 0), \qquad (r \geq 0). \tag{4}$$

Using the limits to $L(\lambda)$, we then find ξ and ξ', which determine En.

2. Special terminology. By an *almost-increasing* function we shall mean one that has the following properties: If x is any point of discontinuity, then (a) $x + k$ is also where k is any integer and $x + l$ is a point of continuity if l is not integral, (b) $f(x - \epsilon) < f(x - \epsilon') < f(x)$ for $0 < \epsilon' < \epsilon < 1$, (c) $f(x - 1) < f(x)$, (d) $\lim_{\epsilon \to 0} f(x + \epsilon) = f(x +) < f(x)$, (e) $f(x - 1 +) < f(x +)$. It is clear that the minimum value for $f(y)$ in any closed interval $[a, b]$ is equal to min $[f(a), f(a' +)]$ where a' is defined as a if the closed interval contains no discontinuity, and as the leftmost point of discontinuity otherwise As special cases, if a is a point of discontinuity this minimum is $f(a +)$ and if $x < a < b < x + 1$ the minimum is $f(a)$.

Almost-decreasing functions are defined similarly except that the inequalities go the other way. In this case the maximum in the interval is $\max[f(a), f(a' +)]$ and we have special cases as above

3. The case $h > 0$. Since $e^z = a^x e^{-c}$, where $a = \lambda_1/\lambda_0$ and $c = (\lambda_1 - \lambda_0)$ the condition $e^{hz} \leq 1/\zeta$ may be expressed as $a^{hx}e^{-ch} \leq 1/\zeta$, whence

$$x \leq c/\log a - \log \zeta/(h \log a) = s - r \text{ (say)}. \tag{5}$$

Since $x \geq 0$, $r \leq s$. Hence $0 < r \leq s$. Also

$$Ee^{zh} = \sum_{x=0}^{\infty} (e^{-c} a^x)^h \frac{e^{-\lambda} \lambda^x}{x!} = \exp(-ch - \lambda + \lambda a^h), \tag{6}$$

and

$$\zeta E(e^{zh} | e^{zh} \leq 1/\zeta) = \zeta E[(e^{-c}a^x)^h | x \leq s - r] \tag{7}$$

From (5), $\zeta = a^{rh}$ and (7) becomes

$$a^{rh} \frac{\sum_{x=0}^{[s-r]} \frac{e^{-\lambda}\lambda^x}{x!} e^{-ch} a^{xh}}{\sum_{x=0}^{[s-r]} \frac{e^{-\lambda}\lambda^x}{x!}}; \tag{7.1}$$

where $[s - r]$ is the largest integer $\leq (s - r)$. Our problem is to minimize (7) with respect to ζ. Since r is a strictly increasing function of ζ, this is equivalent to minimizing $a^{rh}C/D = \theta$ (say) with respect to r, where

$$C = \sum_{x=0}^{[s-r]} \frac{\lambda^x a^{xh}}{x!}, \qquad \text{and} \qquad D = \sum_{x=0}^{[s-r]} \frac{\lambda^x}{x!}.$$

It will be shown that (7 1) is an almost-increasing function of r and therefore the minimum occurs at either $r = 0$ or $r = \nu +$, where $\nu = s - [s]$, since the saltuses occur at $r = \nu + k$ for $k = 0, 1, 2, \cdots, [s]$.

Since a^{rh} is an increasing function of r and C/D remains constant as long as $[s - r]$ remains constant, condition (b) is fulfilled.

Conditions (c) to (e) refer to the saltuses only, hence, to show them, we may assume, without loss of generality that r and s are integral. We proceed by induction, using the notation $\theta(w)$ to mean the value of θ, when $r = w$, to show (c).

First we prove the following:

Lemma A $\theta(s) > \theta(s - 1)$.

Proof: Since we assumed $\lambda_1 > \lambda_0$ and $h > 0$, $a^h > 1$. Hence $(1 + \lambda)a^h > 1 + \lambda a^h$, whence, *a fortiori*, $a^{sh} > a^{(s-1)h}(1 + \lambda a^h)/(1 + \lambda)$.

To show that if $\theta(r + 1) > \theta(r)$, then $\theta(r) > \theta(r - 1)$, we shall show that

$$CD + Dba^{(n+1)h} < CDa^h + Cb \tag{8}$$

implies

$$CD + Dbqa^{(n+1)h} < CDa^h + Cbqa^h, \tag{9}$$

where $n = s - r$, $b = \lambda^n/n!$, $q = \lambda/(n + 1)$.

Since, as we shall see below,

$$Dba^{(n+1)h}(q - 1) < Cb(qa^h - 1), \tag{10}$$

or

$$Da^{(n+1)h}(q - 1) < C(qa^h - 1), \tag{11}$$

addition of (8) and (10) yields the desired result, (9).

It now remains to prove (11) or that

$$\left[\sum_{x=0}^{n} \frac{\lambda^x}{x!}\right] a^{(n+1)h} (\lambda - n - 1) < \left[\sum_{x=0}^{n} \frac{\lambda^x a^{xh}}{x!}\right] (\lambda a^h - n - 1). \tag{12}$$

Setting (6) equal to 1 we get $\lambda a^h = ch + \lambda$, which when substituted in (12) yields

$$(ch + \lambda)^{n+1}(\lambda - n - 1) \sum_{x=0}^{n} \frac{\lambda^x}{x!} < \lambda^{n+1}(ch + \lambda - n - 1) \sum_{x=0}^{n} \frac{(ch + \lambda)^x}{x!} .$$

Upon letting $p = ch + \lambda$, we have

$$\frac{\lambda - (n + 1)}{\lambda^{n+1}} \sum_{x=0}^{n} \frac{\lambda^x}{x!} < \frac{p - (n + 1)}{p^{n+1}} \sum_{x=0}^{n} \frac{p^x}{x!} = F(p), \qquad \text{say}$$

Then our problem reduces to showing that $F(y)$ is increasing in $0 < \lambda \leq y \leq p$ or that the derivative with respect to y, $F'(y)$ is positive.

$$F'(y) = - \sum_{x=0}^{n-1} \frac{(n - x)(y^{x-n-1})}{x!} + (n + 1) \sum_{x=0}^{n-1} \frac{(n - x)(y^{x-n-1})}{(x + 1)!} + (n + 1)^2 y^{-n-2}$$

$$> (n + 1)^2 y^{-n-2}, \qquad \text{since } (n + 1) > (x + 1);$$

$$> 0 \qquad \text{since } y > 0.$$

Thus condition (c) is demonstrated. To show (d) we must show that $\theta(r +) < \theta(r)$, which means that

$$a^{rh} \frac{C - ba^{nh}}{D - b} < a^{rh} \frac{C}{D} .$$

But this is true if $C < Da^{nh}$ which is easily verfied. Condition (e) is equivalent to showing that

$$a^{(r-1)h} \frac{C}{D} < a^{rh} \frac{C - ba^{nh}}{D - b} ,$$

which is proved just as (c) was.

Hence,

$$\eta = \min \left\{ e^{-ch} \sum_{x=0}^{[s]} \frac{e^{-\lambda} \lambda^x a^{hx}}{x!} \Big/ \sum_{x=0}^{[s]} \frac{e^{-\lambda} \lambda^x}{x!} , \right. \tag{13}$$
$$\left. a^{\nu h} e^{-ch} \sum_{x=0}^{[s-1]} \frac{e^{-\lambda} \lambda^x a^{hx}}{x!} \Big/ \sum_{x=0}^{[s-1]} \frac{e^{-\lambda} \lambda^x}{x!} \right\}.$$

As special cases we have (i) if s is integral, η is the latter with $\nu = 0$ and (ii) if $s < 1$ (b) is the only applicable condition and we have an ordinary increasing function, hence η is the former.

Similarly, it may be shown that

$$\delta = \max [e^{-ch} E(a^{xh} | x \geq \{s\}), \quad a^{-\mu h} e^{-ch} E(a^{xh} | x \geq \{s + 1\})], \tag{14}$$

where $\{s\}$ is the smallest integer $\geq s$ and $\mu = \{s\} - s$. Here there is only one special case, namely (i). If $h < 0$, δ is the larger of the two expressions on the right side of (13) and η is the smaller of the two corresponding expressions in (14).

4. Since $z = -c + x \log a$, ξ may be written

$$\operatorname*{Max}_t \log aE(x - t \mid x \geq t),$$

where $t = (r + c)/(\log a)$. Hence $s = c/\log a \leq t < \infty$. Therefore if we can show that $E(x - t \mid x \geq t) = \gamma(t)$ (say), is an almost-decreasing function of t we will know that ξ occurs either when $t = s$ or $\{s\}+$ since, as will be seen, the jumps occur at integral t

To show (c) we make use of the following which is easily proven:

LEMMA B *Let X, Y, Z each be greater than zero. Then a necessary and sufficient condition that $\dfrac{X}{Y} < \dfrac{X+Y}{Y+Z}$ is that $XZ < Y^2$.*

Therefore, to show for integral t that

$$\gamma(t) < \gamma(t-1), \tag{15}$$

or that

$$\frac{\sum_{x=t}^{\infty} (x-t)\frac{\lambda^x}{x!}}{\sum_{x=t}^{\infty}\frac{\lambda^x}{x!}} < \frac{\sum_{x=t}^{\infty}(x-t)\frac{\lambda^x}{x!} + \sum_{x=t}^{\infty}\frac{\lambda^x}{x!}}{\sum_{x=t}^{\infty}\frac{\lambda^x}{x!} + \frac{\lambda^{t-1}}{(t-1)!}},$$

we need only show that, for all integral t,

$$\frac{\lambda^{t-1}}{(t-1)!}\sum_{x=t}^{\infty}\frac{(x-t)\lambda^x}{x!} < \left[\sum_{x=t}^{\infty}\frac{\lambda^x}{x!}\right]^2. \tag{16}$$

Since both sides of (16) are power series in λ where the exponents start with $2t$ we need only show that the coefficient of every term on the left is less than the corresponding term on the right.

In the case of the coefficient of λ^{2j+2t}, $(j \geq 0)$ we have to show that

$$\frac{2j+1}{(t+2j+1)!(t-1)!} < \frac{2}{(t+2j)!t!} + \frac{2}{(t+2j-1)!(t+1)!} + \cdots + \frac{1}{(t+j)!(t+j)!},$$

or by multiplying both sides by $(2t+2j)!$ that

$$(2j+1)\binom{2t+2j}{t-1} < 2\binom{2t+2j}{t} + 2\binom{2t+2j}{t+1} + \cdots + 2\binom{2t+2j}{t+j-1} + \binom{2t+2j}{t+j} = M, \text{ say}$$

Replacing all the binomial coefficients on the right by the smallest one we have

$$(2j+1)\binom{2t+2j}{t-1} < (2j+1)\binom{2t+2j}{t} < M,$$

since $\binom{n}{s-1} < \binom{n}{s}$ for $n \geq 2s$ Thus the truth of (16) has been established for even exponents The odd terms are treated similarly.

Hence, we have shown that $\gamma(t)$ is a strictly decreasing function of t, if t takes on integral values only. We shall now show (b), i.e. that

$$\gamma(t) = \frac{\sum_{x=t}^{\infty} (x-t)\frac{\lambda^x}{x!}}{\sum_{x=t}^{\infty} \frac{\lambda^x}{x!}} < \frac{\sum_{x=[t-\epsilon]}^{\infty} (x-t+\epsilon)\frac{\lambda^x}{x!}}{\sum_{x=[t-\epsilon]}^{\infty} \frac{\lambda^x}{x!}} = \gamma(t-\epsilon). \tag{17}$$

The denominators are equal and each term of the numerator on the right is greater than the corresponding term on the left, hence (17) is valid.

Conditions (a) and (d) can be shown, by showing in a similar manner, that

$$\gamma(t+) = 1 + \gamma(t+1) \tag{18}$$

and $\gamma(t) > 1 + \gamma(t+1)$ for integral t. By using (18) for t and $t-1$ together with (15) we show $\gamma(t-1+) < \gamma(t+)$, which is condition (e). Thus we have shown that

$$\xi = \max\begin{cases} -c + \log a \sum_{x=[s]}^{\infty} \dfrac{x\lambda^x e^{-\lambda}}{x!} \Big/ \sum_{x=[s]}^{\infty} \dfrac{\lambda^x e^{-\lambda}}{x!}, \\ \log a\left[-\{s\} + \sum_{x=[s+1]}^{\infty} \dfrac{x\lambda^x e^{-\lambda}}{x!} \Big/ \sum_{x=[s+1]}^{\infty} \dfrac{\lambda^x e^{-\lambda}}{x!}\right]. \end{cases}$$

As in Section 3, ξ' is the lower analogue of ξ, i.e.

$$\xi' = \min\{-c + E(x \mid x \leq [s]), -[s]\log a + E(x \mid x \leq [s-1])\},$$

and the special cases are as in that section.

REFERENCES

[1] A WALD, "On cumulative sums of random variables," *Annals of Math. Stat.*, Vol. 15 (1944), pp 283–296

[2] A WALD, "Sequential tests of statistical hypotheses," *Annals of Math. Stat.*, Vol. 16 (1945), pp 117–186

NOTES

This section is devoted to brief research and expository articles and other short items.

THE DISTRIBUTION OF STUDENT'S t WHEN THE POPULATION MEANS ARE UNEQUAL

By Herbert Robbins

Department of Mathematical Statistics, University of North Carolina

Let $x_1, \cdots, x_N$ be independent normal variates with the same variance σ^2 and with means $\mu_1, \cdots, \mu_N$ respectively. Set $n = N - 1$ and let

$$(1) \qquad \bar{x} = \sum_1^N x_i/N, \qquad s^2 = \sum_1^N (x_i - \bar{x})^2/n, \qquad t = N^{\frac{1}{2}}\bar{x}/s.$$

If all the μ_i are 0 then t has Student's distribution with n degrees of freedom; its frequency function will be denoted here by

$$(2) \qquad f_{n,0}(t) = n^{-\frac{1}{2}}\left[B\left(\frac{1}{2}, \frac{n}{2}\right)\right]^{-1} \cdot (1 + t^2/n)^{-\frac{1}{2}(n+1)}.$$

When dealing with situations involving mixtures of populations or in which the mean exhibits a secular trend, it is important to know the distribution of t when the μ_i are arbitrary; in the general case let

$$(3) \qquad \begin{aligned} \bar{\mu} &= \sum_1^N \mu_i/N, \qquad \beta^2 = \sum_1^N (\mu_i - \bar{\mu})^2/N, \\ \alpha &= N\bar{\mu}^2/2\sigma^2, \qquad \lambda = N\beta^2/2\sigma^2. \end{aligned}$$

The distribution of t will be shown to depend on the three parameters n, α, λ If $\lambda = \beta^2 = 0$, so that all the μ_i are equal, then the distribution of t determines the power function of the ordinary t test. We shall here consider the case in which $\alpha = \bar{\mu} = 0$, although the μ_i are different. Denoting the frequency function of t in this case by $f_{n,\lambda}(t)$ we shall show that

$$(4) \qquad f_{n,\lambda}(t) = f_{n,0}(t) \cdot \exp\left\{\frac{-\lambda t^2/n}{1 + t^2/n}\right\} \cdot F(-\tfrac{1}{2}, n/2, -\lambda(1 + t^2/n)^{-1}),$$

where F denotes the confluent hypergeometric series, and where, since $\bar{\mu} = 0$,

$$(5) \qquad \lambda = \sum_1^N \mu_i^2/2\sigma^2$$

In fact, the general distribution of t, of which (4) represents the case $\alpha = 0$,

may be derived as follows. Using the standard orthogonal transformation [1, p. 387] let

(6) $$z_i = \sum_{j=1}^{N} c_{ij} x_j, \qquad x_i = \sum_{j=1}^{N} c_{ji} z_j \qquad (i = 1, \cdots, N),$$

where

(7) $$c_{1j} = N^{-\frac{1}{2}} \qquad (j = 1, \cdots, N);$$

then

(8) $$t = n^{\frac{1}{2}} z_1 \Big/ \left(\sum_{2}^{N} z_i^2 \right)^{\frac{1}{2}}.$$

The joint frequency function of the z_i is easily seen to be

(9) $$(2\pi)^{-N/2} \cdot \sigma^{-N} \cdot \exp\left\{ -\sum_{1}^{N} (z_i - a_i)^2 / 2\sigma^2 \right\},$$

where

(10) $$a_1 = N^{\frac{1}{2}} \bar{\mu}, \qquad \sum_{2}^{N} a_i^2 = N\beta^2.$$

Thus t is the ratio of a non-central normal variate to the square root of an independent non-central chi-square variate. It is known [2, p. 138] that the frequency function of $q^2 = \sum_{2}^{N} z_i^2/\sigma^2$ is

(11) $$\tfrac{1}{2} e^{-\lambda} \cdot (\tfrac{1}{2} q^2)^{\frac{1}{2}n - 1} \cdot e^{-q^2/2} \cdot \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda q^2)^j}{j!\,\Gamma(\frac{1}{2}n + j)},$$

where

(12) $$\lambda = \sum_{2}^{N} a_i^2 / 2\sigma^2 = N\beta^2 / 2\sigma^2.$$

The frequency function of $v = z_1/\sigma$ is

$$g(v) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{(\sigma v - a_1)^2}{2\sigma^2} \right\} = \frac{1}{\sqrt{2\pi}} e^{-\alpha} \cdot e^{-(v^2/2)} \cdot \sum_{k=0}^{\infty} \frac{(2\alpha)^{k/2}}{k!} x^k,$$

that of q is, by (11),

$$h(q) = 2^{1-(n/2)} e^{-\lambda} e^{-(q^2/2)} \sum_{j=0}^{\infty} \frac{\lambda^j q^{2j+n-1}}{2^j j!\,\Gamma((n/2)+j)}, \qquad (q > 0),$$

hence that of $u = v/q = n^{-\frac{1}{2}} t$ is

$$\int_0^\infty h(q) g(uq) q \, dq,$$

which, after integration, reduces to

(13) $$\pi^{-\frac{1}{2}} e^{-(\lambda+\alpha)} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{\lambda^j (2\alpha^{\frac{1}{2}} u)^k}{j!\,k!} \frac{\Gamma(N/2 + j + k/2)}{\Gamma(n/2 + j)} (1 + u^2)^{-(N+2j+k)/2}.$$

In particular, if $\alpha = \bar{\mu} = 0$ then (13) reduces by means of the relation $F(\alpha, \gamma, x) = e^x F(\gamma - \alpha, \gamma, -x)$ to

$$\text{(14)} \quad \left[B\left(\frac{1}{2}, \frac{n}{2}\right)\right]^{-1} \cdot e^{-\lambda u^2/(1+u^2)} \quad (1 + u^2)^{-\frac{1}{2}N} \cdot F\left(-\frac{1}{2}, \frac{n}{2}, -\lambda(1 + u^2)^{-1}\right),$$

from which it follows that the frequency function of t is given by (4).

Again, let $x_1, \cdots, x_{N_1+N_2}$ be independent normal variates with the same variance σ^2 and with means $\mu_1, \cdots, \mu_{N_1+N_2}$ respectively. Set $n_1 = N_1 - 1$, $n_2 = N_2 - 1$, $n = n_1 + n_2$, and let

$$\bar{x}_1 = \sum_1^{N_1} x_i/N_1, \qquad \bar{x}_2 = \sum_{N_1+1}^{N_1+N_2} x_i/N_2$$

$$\text{(15)} \qquad s_1^2 = \sum_1^{N_1} (x_i - \bar{x}_1)^2/n_1, \qquad s_2^2 = \sum_{N_1+1}^{N_1+N_2} (x_i - \bar{x}_2)^2/n_2$$

$$s^2 = (n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2), \qquad t = [N_1 N_2/(N_1 + N_2)]^{\frac{1}{2}}(\bar{x}_1 - \bar{x}_2)/s.$$

If all the μ_i are equal then t again has Student's distribution with n degrees of freedom. In the general case let

$$\text{(16)} \qquad \begin{aligned} \bar{\mu}_1 &= \sum_1^{N_1} \mu_i/N_1, & \bar{\mu}_2 &= \sum_{N_1+1}^{N_1+N_2} \mu_i/N_2, \\ \beta_1^2 &= \sum_1^{N_1} (\mu_i - \bar{\mu}_1)^2/N_1, & \beta_2^2 &= \sum_{N_1+1}^{N_1+N_2} (\mu_i - \bar{\mu}_2)^2/N_2. \end{aligned}$$

Then we may show as before [1, p. 388] that in this case $u = n^{-\frac{1}{2}}t$ has the frequency function (13), where now

$$\text{(17)} \qquad \begin{aligned} N &= N_1 + N_2 - 1, \qquad \lambda = (N_1\beta_1^2 + N_2\beta_2^2)/2\sigma^2, \\ \alpha &= [N_1 N_2/(N_1 + N_2)](\bar{\mu}_1 - \bar{\mu}_2)^2/\sigma^2. \end{aligned}$$

In particular, when $\alpha = \bar{\mu}_1 - \bar{\mu}_2 = 0$, so that $\bar{\mu}_1 = \bar{\mu}_2 = \bar{\mu}$, say, the frequency function $f_{n,\lambda}(t)$ of t is again given by (4), where now

$$\text{(18)} \qquad \lambda = \sum_1^{N_1+N_2} (\mu_i - \bar{\mu})^2/2\sigma^2$$

Extensions in this direction to the general linear hypothesis in the analysis of variance will not be treated here

If we set

$$\text{(19)} \qquad w = (1 + t^2/n)^{-1}$$

where t has the frequency function (4), then w will have the frequency function

$$\text{(20)} \quad g_{n,\lambda}(w) = \left[B\left(\frac{1}{2}, \frac{n}{2}\right)\right]^{-1} \cdot e^{-\lambda(1-w)} \cdot w^{\frac{1}{2}n-1} \quad (1 - w)^{-\frac{1}{2}} \cdot F\left(-\frac{1}{2}, \frac{n}{2}, -\lambda w\right),$$

for $0 < w \leq 1$. Thus for every t,

$$(21) \qquad 1 - \int_{-t}^{t} f_{n,\lambda}(x)\, dx = \int_0^{(1+t^2/n)^{-1}} g_{n,\lambda}(w)\, dw.$$

It would be interesting to have numerical values of the integral on the left side of (21) for that value of t for which

$$(22) \qquad 1 - \int_{-t}^{t} f_{n,0}(x)\, dx = 0.01 \text{ or } 0.05 \quad \text{(say)},$$

but existing tables (e.g. those in [2] and [3]) of the integral of (20) were compiled for a different purpose and do not supply this information. The following remarks throw some light on this subject

Let us set

$$(23) \qquad \begin{aligned} R(t) = f_{n,\lambda}(t)/f_{n,0}(t) &= \exp\left\{\frac{-\lambda t^2/n}{1 + t^2/n}\right\} \cdot F\left(-\frac{1}{2}, \frac{n}{2}, -\lambda(1 + t^2/n)^{-1}\right) \\ &= \{1 - \lambda(t^2/n)/(1 + t^2/n) + o(\lambda)\} \\ &\qquad \cdot \{1 + \lambda/(n + t^2) + o(\lambda)\} \\ &= 1 + \lambda(n + t^2)^{-1}(1 - t^2) + o(\lambda). \end{aligned}$$

Then as $\lambda \to 0$ we have ultimately

$$(24) \qquad \begin{aligned} &R(t) > 1 \text{ if } |t| < 1, \\ &R(t) < 1 \text{ if } |t| > 1 \end{aligned}$$

Hence for any $t > 1$ and for sufficiently small λ,

$$(25) \qquad 1 - \int_{-t}^{t} f_{n,\lambda}(x)\, dx < 1 - \int_{-t}^{t} f_{n,0}(x)\, dx.$$

The exact range of values of t for which $R(t) < 1$ depends of course on n and λ. However we shall show that always

$$(26) \qquad R(t) < 1 \text{ if } |t| > 1,$$

so that (25) holds for all n and $\lambda > 0$, provided $t > 1$ The proof is as follows. In terms of w we have

$$(27) \qquad R(t) = e^{-\lambda(1-w)}\, F(-\tfrac{1}{2}, n/2, -\lambda w) = e^{-\lambda} F((n + 1)/2, n/2, \lambda w).$$

Now

$$(28) \qquad \begin{aligned} F((n + 1)/2, n/2, \lambda w) = 1 \\ + \sum_{k=1}^{\infty} \frac{(n + 1)(n + 3) \cdots (n + 2k - 1)}{n(n + 2) \cdots (n + 2k - 2)} (\lambda w)^k/k!, \end{aligned}$$

and by induction on k we may show that for all $k = 1, 2, \cdots$,

$$\frac{(n+1)(n+3)\cdots(n+2k-1)}{n(n+2)\cdots(n+2k-2)} \leq 1 + k/n, \tag{29}$$

where the equality holds only for $k = 1$. Hence

$$F((n+1)/2, n/2, \lambda w) < 1 + \sum_{k=1}^{\infty} (1 + k/n)\cdot(\lambda w)^k/k! = e^{\lambda w}(1 + \lambda w/n), \tag{30}$$

$$R(t) < e^{-\lambda(1-w)}\cdot(1 + \lambda w/n) < e^{-\lambda(1-w)}\cdot e^{\lambda w/n} = e^{-\lambda[1-w(1+1/n)]}. \tag{31}$$

Hence $R(t) < 1$ if $w < n/(n+1)$, which is equivalent to (26).

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946

[2] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat Res Memoirs*, Vol. 2 (1938), pp. 127–149.

[3] Emma Lehmer, "Inverse tables of probabilities of errors of the second kind," *Annals of Math Stat*, Vol 15 (1944), pp 388–398.

A DISTRIBUTION-FREE CONFIDENCE INTERVAL FOR THE MEAN

By Louis Guttman

Cornell University

1. Summary. Consider a random sample of N observations $x_1, x_2, \cdots, x_N$, from a universe of mean μ and variance σ^2 Let m and s^2 be the sample mean and variance respectively:

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad s^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2 \tag{1}$$

It is shown that the following conservative confidence interval holds for μ:

$$\text{Prob}\,\{(m-\mu)^2 \leqq s^2/(N-1) + \lambda\sigma^2\sqrt{2/N(N-1)}\} > 1 - \lambda^{-2}, \tag{2}$$

where λ is any positive constant Inequality (2) also holds if, in the braces, λ is replaced by $\sqrt{\lambda^2 - 1}$, with $\lambda \geqq 1$.

Inequality (2) is much more efficient on the average than Tchebychef's inequality for the mean, namely,

$$\text{Prob}\,\{(m-\mu)^2 \leqq \lambda^2\sigma^2/N\} > 1 - \lambda^{-2}, \tag{3}$$

yet (2) and (3) are both distribution-free, requiring only knowledge about σ^2. At the $1 - \lambda^{-2} = .99$ level of confidence, the expected value of the right member in the braces of (2) is only about 1/6 the corresponding member of (3); at the .999 level of confidence the ratio is about 1/20.

A more general inequality than (2) is developed, also involving only the single parameter σ^2.

2. Derivation. Consider the function

$$u = (m - \mu)^2 - s^2/(N - 1) - c\sigma^2, \tag{4}$$

where c is an arbitrary constant. It is easily verified that $Eu = -c\sigma^2$, and that

$$\underline{E}u^2 = \sigma^4[2/N(N - 1) + c^2]. \tag{5}$$

A basic feature of (5) is that the only population parameter in the right member is σ^2. Contrary to what might have been surmised, the fourth moment of x about μ is not involved, and indeed need not exist.

According to Tchebychef's inequality,

$$\text{Prob}\,\{-\lambda\sqrt{\underline{E}u^2} \leqq u \leqq \lambda\sqrt{\underline{E}u^2}\} > 1 - \lambda^{-2}, \tag{6}$$

where λ is an arbitrary positive number. Using (4) and (5), it is possible to write (6) as:

$$\begin{aligned} \text{Prob}\,\{s^2/(N - 1) + c\sigma^2 - \lambda\sigma^2\sqrt{2/N(N - 1) + c^2} &\leqq (m - \mu)^2 \\ &\leqq s^2/(N - 1) + \sigma^2[c + \lambda\sqrt{2/N(N - 1) + c^2}]\} > 1 - \lambda^{-2}. \end{aligned} \tag{7}$$

In the braces of (7), if the left member is negative, there is no harm in replacing it by zero, if it is positive, then replacing it by zero may only increase the probability of the braces. Regardless of the value of this left member, it is true that

$$\begin{aligned} \text{Prob}\,\{(m - \mu)^2 \leqq s^2/(N - 1) \\ + \sigma^2[c + \lambda\sqrt{2/N(N - 1) + c^2}]\} > 1 - \lambda^{-2}. \end{aligned} \tag{8}$$

If we set $c = 0$, we have inequality (2). Some improvement over (2) is obtained by determining c to minimize the right member in the braces of (8), yielding as the shortest confidence interval:

$$\text{Prob}\,\{(m - \mu)^2 \leqq s^2/(N - 1) + \sigma^2\sqrt{2(\lambda^2 - 1)/N(N - 1)}\} > 1 - \lambda^{-2}. \tag{9}$$

Inequality (9) differs from (2) only by replacing λ in the braces by $\sqrt{\lambda^2 - 1}$.

3. Comparison with Tchebychef's inequality. The expected value of the right member of the braces in (2) is

$$\sigma^2[1/N + \lambda\sqrt{2/N(N - 1)}]. \tag{10}$$

The ratio of (10) to the corresponding value of Tchebychef's inequality (3), namely $\lambda^2\sigma^2/N$, is

$$[1 + \lambda\sqrt{2N/(N - 1)}]/\lambda^2 \tag{11}$$

Since (11) decreases as λ increases, the efficiency of inequality (2) increases compared with that of Tchebychef as the level of confidence $1 - \lambda^{-2}$ increases The

squared interval of (2) involves only the first power of λ, while that of (3) involves the second power

4. Approach to normality. If the fourth moment of the universe's distribution exists, then it is well known that the ratio of $\underline{E}(m - \mu)^4$ to σ^4/N^2 must approach 3—the ratio for the normal distribution—as N increases. That is, if $\alpha^2 + 1$ is the ratio, then $\lim_{N\to\infty} \alpha^2 = 2$ It is known[1] that Tchebychef's inequality can be replaced by one involving both α^2 and σ^2, and that

$$\text{(12)} \qquad \text{Prob }\{(m - \mu)^2 \leqq \sigma^2(1 + \lambda\alpha)/N\} > 1 - \lambda^{-2}.$$

If $\alpha^2 = 2$, then the right member in the braces of (12) becomes $\sigma^2(1 + \lambda\sqrt{2})/N$ This is virtually the same as (10), the expected value from (2). In a sense, then, (2) implicitly takes account of the fact that the distribution of sample means approaches that of the normal distribution with respect to the fourth moment. A striking feature, however, is that (2) holds for any $N > 1$ and does not even presume the fourth moment of the universe to exist, whereas to set $\alpha = \sqrt{2}$ in (12) in general requires a large N and finite universe fourth moment

5. Further possibilities. Confidence interval (2) is derived from but one of a series of general intervals, each of which depends only on σ^2. It may be possible to derive from this series even more efficient intervals, according to the method now to be outlined

One way of arriving at (2) is to consider all products of the form $(x_i - \mu)(x_j - \mu)$, where $i > j$ and $i, j = 1, 2, \cdots, N$. Let p_2 be the mean of these $N(N - 1)/2$ products It can easily be seen that $p_2 = u$ in (4) with $c = 0$, so that p_2 is a second degree polynomial in $m - \mu$, the coefficients being sample statistics A more general quadratic would be $u_2 = p_2 + c_1p_1 + c_0$, where c_1 and c_0 are arbitrary constants and p_1 is the mean of the N values $(x_i - \mu)$ or $p_1 = m - \mu$. It is easily seen that $\underline{E}p_1 = \underline{E}p_2 = \underline{E}p_1p_2 = 0$, and that the only universe parameter involved in $\underline{E}p_1^2$ and $\underline{E}p_2^2$ is σ^2. Hence the only universe parameter upon which u_2^2 depends is also σ^2

Higher degree polynomials in $m - \mu$ can be defined, possessing the same properties as u_2 Let p_3 be the mean of the $N(N - 1)(N - 2)/3!$ products of the form $(x_i - \mu)(x_j - \mu)(x_k - \mu)$, where $i > j > k$ and $i, j, k = 1, 2, \cdots, N$; etc., and let $p_N = (x_1 - \mu)(x_2 - \mu) \cdots (x_N - \mu)$. Set $p_0 = 1$, and let

$$\text{(13)} \qquad u_n = \sum_{a=0}^{n} c_{an}\, p_a \qquad (n = 1, 2, \cdots, N),$$

where the c_{an} are arbitrary constants. It is easily seen that $\underline{E}p_a = 0$ $(a > 0)$, $\underline{E}p_ap_b = 0$ $(a \neq b)$, and that each $\underline{E}p_a^2$ depends on only the parameter σ^2 as far

[1] See, for example, Louis Guttman, "An inequality for kurtosis," *Annals of Math Stat.*, Vol. 19 (1948), pp. 277–278.

as the universe is concerned. Hence $\underline{E}u_n^2$ depends only on σ^2 Furthermore, by writing $x_i - \mu$ as $(x_i - m) + (m - \mu)$, it is seen that p_a is a polynomial of degree a in $m - \mu$, the coefficients being sample statistics. From (13), then, u_n is a polynomial of degree n in $m - \mu$ with statistics as coefficients

According to Tchebychef's inequality,

$$\text{Prob } \{u_n^2 \leqq \lambda^2 \underline{E}u_n^2\} > 1 - \lambda^{-2} \tag{14}$$

The interval for u_n^2 in the braces can be expressed in two statements:

$$f_n(m - \mu) = u_n - \lambda\sqrt{\underline{E}u_n^2} \leqq 0, \tag{15}$$

$$g_n(m - \mu) = u_n + \lambda\sqrt{\underline{E}u_n^2} \geqq 0 \tag{16}$$

Both f_n and g_n are polynomials of degree n in $m - \mu$, g_n exceeding f_n always by the additive constant $2\lambda\sqrt{\underline{E}u_n^2}$ Let q_n and Q_n be the smallest and largest real zeros respectively of f_n, and let r_n and R_n be the smallest and largest real zeros respectively of g_n

For convenience, we can suppose that c_{nn}—the coefficient of $(m - \mu)^n$ in u_n—is positive. If n is even, then f_n is positive for $m - \mu > Q_n$ and for $m - \mu < q_n$. Hence the interval $q_n \leqq m - \mu \leqq Q_n$ contains all the points included in (15) and possibly more. Since the probability of (15) is not less than the probability of (14), we can write the following confidence interval:

$$\text{Prob } \{q_n \leqq m - \mu \leqq Q_n\} > 1 - \lambda^{-2} \qquad (n \text{ even}). \tag{17}$$

The problem remains to determine the c_{an} so as to minimize the expected value of $Q_n - q_n$ Inequality (9) provides the minimum for the case $n = 2$ This can be verified by adding the term $c_1 p_1$ to u in (4) and finding that the minimum requires $c_1 = 0$.

If n is odd, we again may set $c_{nn} > 0$. Then $f_n > 0$ for $m - \mu > Q_n$, and $g_n < 0$ for $m - \mu < r_n$. The interval $r_n \leqq m - \mu \leqq Q_n$ thus contains at least all the points found jointly in (15) and (16) and hence forms a conservative confidence interval:

$$\text{Prob } \{r_n \leqq m - \mu \leqq Q_n\} > 1 - \lambda^{-2} \qquad (n \text{ odd}) \tag{18}$$

Again, the problem is to determine the c_{an} so as to minimize the expected value of $Q_n - r_n$ Tchebychef's inequality (3) does this for the case $n = 1$

Although the only *population* parameter involved throughout is σ^2, the *sample* moments up to the nth order are present in (15) and (16) It thus seems plausible that improvement over inequality (9) should be possible for $n > 2$ To obtain such an improvement requires developing a distribution-free theory of the zeros of f_n and g_n beyond the quadratic case

ON THE COMPOUND AND GENERALIZED POISSON DISTRIBUTIONS

By E. Cansado Maceda

University of Madrid

1. Summary. In this note we deduce several properties of the compound and generalized Poisson distributions, in particular their closure and divisibility properties. An infinite class of functions whose members are both compound and generalized Poisson distributions is exhibited, and several of the distributions of Neyman, Polya, etc. are identified The present note stems from a paper by Feller [2]

2. The compound Poisson distribution. If $F(x \mid a)$ is a family of distribution functions depending on the parameter a, and $U(a)$ is a distribution function such that it assigns zero probability to any a domain for which $F(x|a)$ is undefined, then

$$G(x) = \int_{-\infty}^{\infty} F(x \mid a) \, dU(a)$$

is a distribution function In particular if $F(x|a)$ is the Poisson distribution with mean a, and $U(0) = 0$, $G(x)$ is called the *compound Poisson distribution* associated with the distribution function $U(a)$, cf Feller [2]. Clearly $G(x)$ is a step function over the non-negative integers, the saltus at the point $x = n$ being

$$\pi_n = \int_0^{\infty} e^{-a} \frac{a^n}{n!} \, dU(a), \qquad n = 0, 1, 2, \cdots$$

It is convenient to introduce the factorial moment generating function (f m.g f.) for $G(x)$ as follows

$$\begin{aligned} \omega(z) &= E((1 + z)^x) = \sum_{n=0}^{\infty} \pi_n (1 + z)^n \\ &= \int_0^{\infty} e^{+az} \, dU(a) \\ &= \phi(z) \end{aligned}$$

where $\phi(z)$ is the ordinary moment generating function (m g f.) for $U(a)$ This gives a convenient relationship between the moments of $U(a)$ and its associated compound Poisson distribution.

On account of the multiplicative properties of $\omega(z)$ and $\phi(z)$ under the convolution of $G(x)$ and $U(a)$ respectively, it is seen that the compound Poisson distributions form a closed family, and if $G_1(x)$ and $G_2(x)$ are two compound Poisson distributions associated with $U_1(a)$ and $U_2(a)$ respectively then $G_1(x) * G_2(x)$ is associated with $U_1(a) * U_2(a)$. In addition, if $U(a)$ is infinitely divisible (cf. Cramér [1]) then $G(x)$ is also, since it can be factored into the convolution of arbitrarily many compound Poisson distributions.

Choosing in particular $U(a)$ as the Pearson type III distribution, the associated function is the Polya-Eggenberger distribution, and if $U(a)$ is a Poisson distribution the associated function is the Neyman contagious distribution of Type A.

3. The generalized Poisson distribution. If $F(x \mid a)$, defined for non-negative integers $a = 0, 1, 2, \cdots$, is the a-fold convolution of a given distribution $F(x)$ with itself, i e. $F(x|a) = F(x)^{*a}$, and $U(a)$ is the Poisson distribution with parameter α, then the distribution function

$$G(x) = \int_0^\infty F(x \mid a)\, dU(a)$$

is called the *generalized Poisson distribution* associated with $F(x)$.

If $\Omega(z)$ is the f.m.g.f of $U(a)$ then for the f m g f of $G(x)$ we have

$$\begin{aligned}\omega(z) &= \sum_{n=0}^{\infty} (\Omega(z))^n e^{-\alpha} \frac{\alpha^n}{n!} \\ &= e^{\alpha(\Omega(z)-1)}.\end{aligned}$$

It follows that $\omega(z)$ can be written as $\prod_{\nu=1}^{\infty} \omega_\nu(z)$ where $\omega_\nu(z)$ is a generalized Poisson distribution, and thus $\omega(z)$ belongs to the infinitely divisible family. Moreover, if $G_1(x)$ and $G_2(x)$ are two generalized Poisson distributions associated with $U_1(a)$ and $U_2(a)$ with parameters α_1 and α_2 respectively, then $G(x) = G_1(x)_* G_2(x)$ has for f.m.g.f

$$\omega_1(z)\omega_2(z) = \exp\left\{(\alpha_1 + \alpha_2)\left(\frac{\alpha_1 \Omega_1(z) + \alpha_2 \Omega_2(z)}{\alpha_1 + \alpha_2} - 1\right)\right\},$$

and $G(x)$ is again a generalized Poisson distribution function associated with the distribution

$$U(a) = \frac{\alpha_1 U_1(a) + \alpha_2 U_2(a)}{\alpha_1 + \alpha_2}$$

and with the parameter $\alpha_1 + \alpha_2$ Thus the generalized Poisson distributions form a closed family. The analytic nature of the generalized Poisson distributions have been studied by Hartman and Wintner [3]. As noted by Feller [2] the various Neyman contagious distributions are generalized Poisson distributions.

4. Further remarks. From the above observations it is clear that a necessary and sufficient condition for a distribution to be a compound Poisson distribution is that its f m g f. be of the form

(1) $$\omega_1(z) = \phi(z)$$

where $\phi(z)$ is the ordinary m.g.f. of a non-negative random variable. Likewise a necessary and sufficient condition for $\omega(z)$ to be the f.m.g.f. of a generalized Poisson distribution is that it be of the form

$$\omega_2(z) = e^{\alpha(\Omega(z)-1)}, \qquad \alpha > 0, \tag{2}$$

where $\Omega(z)$ is the f.m.g.f. of an arbitrary distribution function $F(x)$. If we choose $\phi(z) = e^{\alpha(e^{cz}-1)}$ and $\Omega(z) = e^{cz}$, then $\omega_1(z) = \omega_2(z)$, and the distribution whose f.m.g.f. is $\omega_1(z)$ (the Neyman contagious distribution of Type A) is simultaneously a compound and a generalized Poisson distribution (cf. Feller [2]). We now show that there is an infinite class of distributions with this property.

First note that if $\phi(z)$ is the m.g.f. of an arbitrary distribution, then exp $\{\alpha(\phi(z) - 1)\}$ is also the m.g.f. of a d.f., and in fact is the m.g.f. of the generalized Poisson distribution associated with the distribution whose m.g.f. is $\phi(z)$. Now let $\phi(z)$ be the m.g.f. of an arbitrary non-negative random variable, and define

$$\omega(z) = \exp\{\alpha(\phi(z) - 1)\} \qquad \alpha > 0. \tag{3}$$

Then $\omega(z)$ is simultaneously of the forms (1) and (2), since $\phi(z)$ is, by (1), also the f.m.g.f. of a distribution function, i.e. the compound Poisson distribution associated with the distribution whose m.g.f. is $\phi(z)$. However, not every distribution which is both a compound and a generalized Poisson distribution can be generated in this manner. For example, the Polya-Eggenberger distribution is easily shown to be both a generalized and a compound Poisson distribution, yet its f.m.g.f.

$$\omega(z) = (1 - dz)^{-h/d}, \qquad d > 0, h > 0,$$

manifestly is not of the form (3), since this would imply $\phi(iz) = 1 - \dfrac{h}{\alpha d} \log (1 - diz)$ is a characteristic function. But $|\phi(iz)|$ is unbounded as $z \to \pm\infty$ and thus is not the characteristic function of a distribution.

REFERENCES

[1] H. Cramér, "Problems in probability theory," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 165–193.

[2] W. Feller, "On a general class of contagious distributions," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 389–400.

[3] P. Hartman and A. Wintner, "On the infinitesimal generators of integral convolutions," *Am. Jour. of Math.*, Vol. 64 (1942), pp. 272–279.

ON CONFIDENCE LIMITS FOR QUANTILES

By Gottfried E. Noether

Columbia University

In finding confidence limits for quantiles it is usual to determine two order statistics Z_i and Z_j which with a given probability contain the unknown quantile

between them. The values of i and j corresponding to a given confidence coefficient can be determined with the help of the distribution laws of order statistics as is shown, e g., in Wilks [1] The purpose of this note is to determine i and j with the help of a confidence band for the unknown cumulative distribution function.

In what follows we shall always denote the cumulative distribution function (cdf) by $F(x)$, i.e., $F(x) = P\{X \leq x\}$. Then the quantile q_p is determined by

$$F(q_p - 0) \leq p \leq F(q_p) \tag{1}$$

which reduces to

$$F(q_p) = p \tag{1'}$$

if $F(x)$ is continuous Given a sample of size n we can construct the sample cdf $F_n(x)$ defined by $F_n(x) = 1/n$ (number of observations $\leq x$) Confidence coefficients will always be denoted by $1 - \alpha$

Assume that we can construct two step functions $L(x)$ and $U(x)$ parallel to $F_n(x)$ such that for any fixed value x

$$P\{L(x) < F(x) < U(x)\} = 1 - \alpha. \tag{2}$$

We do not require that the confidence band determined by $L(x)$ and $U(x)$ cover the graph of the unknown cdf $F(x)$ with probability $1 - \alpha$, but only that for any arbitrarily chosen value x (2) is true

Let

$$L(x) = \eta_k, \qquad U(x) = \theta_k$$

for $z_k \leq x < z_{k+1}$, $k = 0, 1, \cdot \quad, n$ where z_k is the value taken by the order statistic Z_k and $z_0 = -\infty$, $z_{n+1} = +\infty$ Then if $F(x)$ is continuous it follows from (2) that a confidence interval with confidence coefficient $1 - \alpha$ for q_p is given by

$$Z_i \leq q_p < Z_j \tag{3}$$

where i and j are determined by

$$\theta_{i-1} \leq p, \qquad \theta_i > p \tag{4}$$

$$\eta_{j-1} < p, \qquad \eta_j \geq p \tag{5}$$

It will be noted that (3) represents a half-open interval. However as long as we only admit continuous cdf's the confidence coefficient is not changed if we use

$$Z_i < q_p < Z_j \tag{3'}$$

or

$$Z_i \leq q_p \leq Z_j \tag{3''}$$

instead. This is no longer true if we also admit discontinuous cdf's. Then the confidence coefficient connected with (3') is $\leq 1 - \alpha$, while that connected with

(3″) is $\geq 1 - \alpha$, as follows immediately from consideration of the possible outcomes when (1) is true. This is the same result as that obtained by Scheffé and Tukey [2]

We shall now indicate how η_k and θ_k can be obtained and find their values in a particular case. For any arbitrary value x we can consider $F_n(x)$ as the sample estimate of the unknown parameter $p = F(x)$ of a binomial distribution Clopper and Pearson [3] have discussed how confidence intervals for the unknown parameter of a binomial variate can be found. Thus we can determine η_k and θ_k correspondingly, but as is well known (2) cannot be achieved with probability exactly equal to $1 - \alpha$. We shall have to be satisfied with probability $\geq 1 - \alpha$. Consequently the same will hold true for the confidence coefficient connected with the confidence interval for q_p.

In many cases central confidence intervals seem to be more desirable, at least intuitively, than others. Our method produces such central confidence intervals for the unknown quantile if we use central confidence intervals in the construction of the confidence band. In that case η_k and θ_k are determined by

$$\frac{\alpha}{2} = I_{\eta_k}(k, n - k + 1) \tag{6}$$

$$\frac{\alpha}{2} = I_{1-\theta_k}(n - k, k + 1), \qquad k = 0, 1, \cdots, n \tag{7}$$

except that $\eta_0 = 0$, $\theta_n = 1$ by definition, where

$$I_x(p, q) = \int_0^x t^{p-1}(1 - t)^{q-1}\, dt \Big/ \int_0^1 t^{p-1}(1 - t)^{q-1}\, dt$$

is the incomplete beta function. Scheffé [4] has pointed out how the tables of percentage points of the incomplete beta function by C. M. Thompson, etc. [5] can be used to find η_k and θ_k.

We shall show now that in the case of the median M the solution based on (3)–(7) leads to the same confidence interval as that suggested originally by W. R. Thompson [6]. Thompson found that for $k < n + \frac{1}{2}$

$$P\{Z_k < M < Z_{n-k+1}\} = 1 - 2I_{\frac{1}{2}}(n - k + 1, k) \tag{8}$$

provided the unknown distribution had a continuous cdf. (8) can be used to maximize k under the condition that the righthand side is $\geq 1 - \alpha$.

We shall first show that our method leads to the same kind of a confidence interval, i.e., one with $i = l, j = n - l + 1$. This follows immediately from the fact that by (6) and (7)

$$1 - \theta_l = \eta_{n-l}. \tag{9}$$

For let

$$\theta_{l-1} \leq \tfrac{1}{2} \text{ and } \theta_l > \tfrac{1}{2}, \tag{10}$$

then by (9) $\eta_{n-l} < \frac{1}{2}$ and $\eta_{n-l+1} \geq \frac{1}{2}$.

It remains to be shown that k as determined by (8) equals l. This will be so if we can show that

$$I_{\frac{1}{2}}(n - l + 1, l) \leq \frac{\alpha}{2} < I_{\frac{1}{2}}(n - l, l + 1) \tag{11}$$

Remembering that $I_x(p, q)$ is a monotonically increasing function of x we get with the help of (7) and (10)

$$\frac{\alpha}{2} = I_{1-\theta_{l-1}}(n - l + 1, l) \geq I_{\frac{1}{2}}(n - l + 1, l)$$

and

$$\frac{\alpha}{2} = I_{1-\theta_l}(n - l, l + 1) < I_{\frac{1}{2}}(n - l, l + 1)$$

which proves (11).

In conclusion it may be worth while pointing out that the formula

$$P\{Z_i < q_p < Z_j\} = I_p(i, n - i + 1) - I_p(j, n - j + 1)$$

given, e.g, in Wilks [1] for the continuous case can be obtained by a slight modification of (6).

REFERENCES

[1] S S. Wilks, "Order statistics," *Am Math Soc. Bull*, Vol 54 (1948), pp. 6–50.
[2] H Scheffé and J W Tukey, "Non-parametric estimation I Validation of order statistics," *Annals of Math Stat*, Vol 16 (1945), pp 187–192
[3] C. J. Clopper and E S Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, Vol 26 (1934), pp 404–413.
[4] H. Scheffé, "Note on the use of the tables of percentage points of the incomplete beta function to calculate small sample confidence intervals for binomial p," *Biometrika*, Vol 33 (1944), p. 181
[5] C. M. Thompson, E S. Pearson, L J Comrie, and H. O. Hartley, "Tables of percentage points of the incomplete beta function," *Biometrika*, Vol. 32 (1941), pp. 151–181.
[6] W. R. Thompson, "On confidence ranges for the median and other expectation distributions for populations of unknown distribution form," *Annals of Math Stat*. Vol 7 (1936), pp 122–128.

A LOWER BOUND FOR THE EXPECTED TRAVEL AMONG m RANDOM POINTS

By Eli S. Marks

Bureau of the Census

In connection with cost determinations in sampling problems, it is frequently necessary to determine the amount of travel among m random sample points in an area A lower bound for the expected value of this distance is found to be:

$$\sqrt{\frac{A}{2}}\,\frac{m-1}{\sqrt{m}},$$

where A is the measure of the area from which the m random points are drawn.[1]

If in a finite area S we locate m points at random (see Figure 1), we can trace a continuous path among the m points by starting at some point and connecting the points by line segments. The points can be connected in any order so that the path touches each point only once (unless it intersects itself at one of the random points). We are interested in a lower bound for the expected value of the length of the shortest of the $m!$ possible paths.

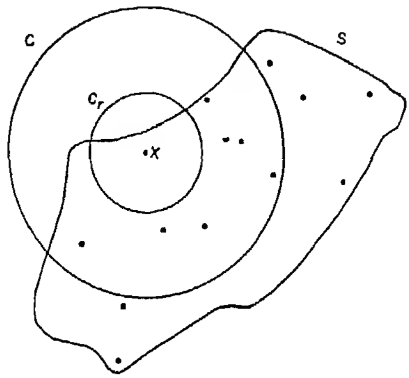


FIG. 1. m RANDOM POINTS IN S.

We have above an area S in which m random points have been selected (with $m = 14$).

The shortest path among the m points consists of $m - 1$ "links" (line segments) between two points. Each link can be assigned to one of its end points, leaving some pre-designated point (e.g., the m-th point selected) with no link assigned. The link assigned to the i-th random point ($x_{(i)}$) must be no less than $r_{(i)}$ the distance from $x_{(i)}$ to the nearest of the other $(m - 1)$ points If we denote the length of the shortest path by L:

$$L \geq \sum_{i=1}^{m-1} r_{(i)},$$

$$E(L) \geq \sum_{i=1}^{m-1} E(r_{(i)})$$

Let $E_x(r_{(i)})$ be the expected value of $r_{(i)}$ conditional upon $x_{(i)}$ falling at the point x in S and let $F(r \mid x)$ be the conditional distribution function of $r_{(i)}$ for $x_{(i)} = x$. Thus $F(r \mid x)$ is the conditional probability of $r_{(i)} \leq r$ or the probability of

[1] The lower bound obtained is similar in form to the expression for distance traveled among a set of random points used by Mahalanobis [2] and Jessen [1]

one or more of the $(m - 1)$ random points other than $x_{(i)}$ falling inside a circle, C_r, with radius r and center at x (see Fig. 1). Then, we have:

$$E_x(r_{(i)}) = \int_0^{+\infty} r\,dF(r|x),$$

$$F(r \mid x) = 1 - \left\{\frac{M(S) - M(SC_r)}{M(S)}\right\}^{m-1},$$

where $M(S)$ and $M(SC_r)$ are the measures of S and SC_r, so that $\dfrac{M(SC_r)}{M(S)}$ is the probability of a random point in S falling into C_r.

Let $A = M(S)$ and construct a circle C with center at x and radius $\rho = \sqrt{\dfrac{A}{\pi}}$. Then $M(C) = A = M(S)$. Let d be the distance from x to the nearest of $(m - 1)$ points selected at random from C and let $G(r)$ be the distribution function of d. Then we have.

$$E(d) = \int_0^{+\infty} r\,dG(r),$$

$$G(r) = 1 - \left\{\frac{M(C) - M(CC_r)}{M(C)}\right\}^{m-1}.$$

For $r \leq \rho$,

$$M(C_rC) = M(C_r) \geq M(SC_r).$$

For $r > \rho$,

$$M(C_rC) = M(C) = M(S) \geq M(SC_r).$$

Thus, since $M(C_rC) \geq M(SC_r)$, we have for all x in S:

$$G(r) \geq F(r \mid x),$$

and thus,

$$E(d) \leq E_x(r_{(i)})$$

Since $E(d) \leq E_x(r_{(i)})$ for all x in S.

$$E(d) \leq E(r_{(i)}),$$

$$(m - 1)E(d) \leq \sum_{i=1}^{m-1} E(r_{(i)}) \leq E(L).$$

It only remains to evaluate $E(d)$, the expected distance from the center of a circle to the nearest of $(m - 1)$ random points This can be done very easily by substituting in the expression for $G(r)$:

$$A = M(C),$$

$$\pi r^2 = M(C_r C), \qquad \text{when } r \leq \rho = \sqrt{\frac{A}{\pi}},$$

to give:

$$G(r) = 1 - \left\{\frac{A - \pi r^2}{A}\right\}^{m-1},$$

$$G'(r) = \frac{2\pi r}{A}(m-1)\left\{\frac{A - \pi r^2}{A}\right\}^{m-2},$$

$$E(d) = \int_0^\rho rG'(r)\,dr = \frac{1}{2}\sqrt{\frac{A}{\pi}}\,[B(m, \tfrac{1}{2})],$$

where $B(m, \frac{1}{2})$ is the complete Beta function.
Since $\sqrt{m}\,[B(m, \frac{1}{2})] \geq \sqrt{\pi}$:

$$E(d) \geq \frac{1}{2}\sqrt{\frac{A}{m}}.$$

Thus, we have:

$$E(L) \geq \tfrac{1}{2}\sqrt{A}\,\frac{m-1}{\sqrt{m}}.$$

It is obvious that the development is general and applies to m random points in any bounded two-dimensional Borel set. However, the lower bound obtained will, in general, be useful only when S is a connected region.

REFERENCES

[1] RAYMOND J. JESSEN, "Statistical investigation of a sample survey for obtaining farm facts," *Iowa State College Research Bulletin* 304 (1942).
[2] P. C. MAHALANOBIS, "A sample survey of the acreage under jute in Bengal," *Sankhyā*, Vol. 4 (1940), pp 511–530

A MATRIX ARISING IN CORRELATION THEORY[1]

BY H. M. BACON

Stanford University

1. Introduction. In the study of time series, it is frequently desirable to consider correlations between observations made in different years. Let x_{i1}, $x_{i2}, \cdots, x_{im}$ be m values of the variable x_i, expressed as deviations from their arithmetic mean, where x_i is a variable observed in the ith year ($i = 1, 2, \cdots, n$).

[1] A linear correlogram is considered by Cochran in his paper, "Relative accuracy of systematic and stratified random samples for a certain class of populations," (*Annals of Math. Stat.*, Vol 17 (1946), pp. 164–177) in which $\rho_\mu = 1 - \frac{\mu}{L}$. Setting $\mu = |i - j|$ and $L = 1/p$, we have the case considered above.

Let σ_i be the standard deviation of x_i. If we denote by $r_{ij} = r_{ji}$ the correlation of x_i with x_j, and if we assume the x_i to be normally distributed, then

$$z = \frac{1}{(2\pi)^{n/2}\sigma_1\sigma_2 \cdots \sigma_n\sqrt{R}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_{ij}\,x_i\,x_j}{R\sigma_i\,\sigma_j}\right\}$$

is the frequency function giving the distribution. Here R is the determinant $|r_{ij}|$ of the correlation coefficients, and R_{ij} is the cofactor of the element r_{ij} in this determinant.

We may make various assumptions regarding the behavior of the correlation coefficients over the n years. One such assumption of some interest is that the correlation coefficients diminish in such a way that

$$r_{ij} = r_{ji} = 1 - |i - j|p$$

where p is a fixed positive number not greater than $2/(n - 1)$. Under these circumstances, we can evaluate R and R_{ij} in terms of n and p.

2. Evaluation of R. We may let $R(p)$ represent the determinant R of order n whose element in the ith row and jth column is $r_{ij} = r_{ji} = r_{n-i,n-j} = r_{n-j,n-i} = 1 - |i - j|p$ where, for the purpose of evaluation, p is any real number. Since each two-rowed minor of $R(p)$ is divisible by p, $R(p)$ is divisible by p^{n-1}. Furthermore, since $R(p)$ is a polynomial in p of degree at most n, we have

$$R(p) = Ap^n + Bp^{n-1} = p^{n-1}(Ap + B).$$

If we set $p = 1$ and $p = -1$, we find $A + B = R(1)$ and $R(-1) = (-1)^{n-1}(-A + B)$ so that $-A + B = (-1)^{n-1}R(-1)$. By elementary methods we find that $R(1) = 2^{n-2}(3 - n)$ and $R(-1) = (-1)^{n-1}2^{n-2}(n + 1)$. Hence

$$A + B = 2^{n-2}(3 - n)$$

and

$$-A + B = 2^{n-2}(n + 1).$$

Solving for A and B we find that

$$R = R(p) = 2^{n-2}p^{n-1}[2 - (n - 1)p].$$

3. Evaluation of R_{ij}. Similar methods yield the following values for the cofactors R_{ij} of the elements of R:

$$\begin{aligned}
R_{11} &= R_{nn} = 2^{n-3}p^{n-2}\,[2 - (n - 2)p],\\
R_{22} &= R_{33} = \cdots = R_{n-1,n-1} = 2^{n-2}p^{n-2}\,[2 - (n - 1)p],\\
R_{1n} &= R_{n1} = 2^{n-3}p^{n-1},\\
R_{i,i+1} &= -2^{n-3}p^{n-2}[2 - (n - 1)p],
\end{aligned}$$

otherwise,

$$R_{ij} = 0.$$

4. The frequency function. The quadratic form appearing in the exponent in the expression for the frequency function can now be written as

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_{ij}\,x_i\,x_j}{R\sigma_i\,\sigma_j} = \frac{2-(n-2)p}{2p[2-(n-1)p]}\left(\frac{x_1^2}{\sigma_1^2}+\frac{x_n^2}{\sigma_n^2}\right)$$

$$+\frac{1}{p}\left(\frac{x_2^2}{\sigma_2^2}+\frac{x_3^2}{\sigma_3^2}+\cdots+\frac{x_{n-1}^2}{\sigma_{n-1}^2}\right)$$

$$+\frac{1}{2[2-(n-1)p]}\left(\frac{x_1\,x_n}{\sigma_1\,\sigma_n}+\frac{x_n\,x_1}{\sigma_n\,\sigma_1}\right)$$

$$-\frac{1}{2p}\left(\frac{x_1\,x_2}{\sigma_1\,\sigma_2}+\frac{x_2\,x_1}{\sigma_2\,\sigma_1}+\frac{x_2\,x_3}{\sigma_2\,\sigma_3}+\frac{x_3\,x_2}{\sigma_3\,\sigma_2}+\cdots+\frac{x_n\,x_{n-1}}{\sigma_n\,\sigma_{n-1}}\right)$$

$$=\frac{1}{p}\left[\frac{2-(n-2)p}{2[2-(n-1)p]}\left(\frac{x_1^2}{\sigma_1^2}+\frac{x_n^2}{\sigma_n^2}\right)+\sum_{i=2}^{n-1}\frac{x_i^2}{\sigma_i^2}-\sum_{i=1}^{n-1}\frac{x_i\,x_{i+1}}{\sigma_i\,\sigma_{i+1}}\right]$$

$$+\frac{1}{2-(n-1)p}\left(\frac{x_1\,x_n}{\sigma_1\,\sigma_n}\right).$$

5. Maximum likelihood. The expression z is the likelihood of getting a particular set of values of the variables $x_1, x_2, \cdots, x_n$. It is often important to regard the r_{ij} and the σ_i as parameters and to determine them so that the likelihood will be a maximum. If we assume $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$, then

$$z = \frac{1}{(2\pi)^{n/2}\sigma^n\sqrt{R}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{R_{ij}\,x_i\,x_j}{R\sigma^2}\right\}.$$

The question, in our case, now becomes, What values of p and σ will make z a maximum for given x_i? Necessary conditions are that $\frac{\partial z}{\partial p} = 0$ and $\frac{\partial z}{\partial \sigma} = 0$. Since R_{ij} and R are given in terms of p, the process of differentiation can be carried out (first take the logarithm of z), and values of p and σ necessary for a maximum determined It is, of course, possible that z has no maximum, and the sufficiency of these values must be tested. The computations for the general case are laborious, though straightforward. Furthermore, because of the complicated nature of the coefficients in the equation to be solved for p, the general solution is not readily obtainable. This equation is, however, of third degree, and it can be solved in any particular case.

TABLE OF NORMAL PROBABILITIES FOR INTERVALS OF VARIOUS LENGTHS AND LOCATIONS

BY W J. DIXON

University of Oregon

1. Introduction. The probability associated with a particular finite range of values is often desired. The usual tables of normal areas gives values for $\int_0^x$ or

as in the table by Salvosa [1], $\int_{-\infty}^{x}$ The WPA table [2] gives $\int_{-x}^{x}$. The author has deposited with Brown University a table of $\int_{x-\frac{1}{2}l}^{x+\frac{1}{2}l}$ for values of $x[0(.1)5.0]$ and values of $l[0(.1)10.0]$. The values in the table may be interpreted as the probability that an observation from a normal population with unit variance will fall in an interval of length l whose midpoint is a distance x from the mean. These values can be obtained by a simple computation from the existing tables. Since values were being used frequently, the present table was constructed. Microfilm or photostat copies may be obtained upon request to the Brown University Library

2. Computation. The values were obtained by finding the difference between the integrals $\int_{-\infty}^{x-\frac{1}{2}l}$ and $\int_{-\infty}^{x+\frac{1}{2}l}$ as given to six decimal places in Salvosa's table. Being differences, the values are subject to an error of 1 unit in the sixth place. For values of $x + \frac{1}{2}l$ greater than 5, the values can be obtained by computing $1 - \int_{-\infty}^{x-\frac{1}{2}l}$. The search for errors was aided by computing column sums, i e

$$\sum_{i=1}^{50} \int_{x_i-\frac{1}{2}l}^{x_i+\frac{1}{2}l} + \frac{1}{2}\int_{-\frac{1}{2}l}^{\frac{1}{2}l} = .5\, n, \tag{1}$$

where i represents the row number and n represents the column number. For example, $n = 17$ corresponds to column for $l = 1.7$ The approximation becomes poorer as n increases but the sums were still useful for checking purposes.

3. Example. The table has been used in studies of the expected proportion of a line covered by intervals dropped on it according to some normal probability function. Let $P_n(x)$ be the probability that the point x is covered at least once when n intervals are dropped on the x-axis H E Robbins [3] gives the expression:

$$E(F) = \frac{1}{L}\int_0^L P_n(x)\, dx, \tag{2}$$

for the expected proportion of a line of length L covered at least once by these intervals

Let $f(x)\, dx$ be the probability that an interval falls with its center in dx and l be the length of the interval The probability that a point x will be covered by one interval dropped on the x-axis is:

$$g(x) = \int_{x-\frac{1}{2}l}^{x+\frac{1}{2}l} f(t)\, dt \tag{3}$$

When n intervals are dropped, the probability that x is covered at least once is·

$$P_n(x) = 1 - (1 - g(x))^n, \tag{4}$$

and

$$E(F) = 1 - \frac{1}{L}\int_0^L (1 - g(x))^n \, dx. \tag{5}$$

When k groups of n_i intervals are dropped according to, say normal distributions with different means,

$$P_n(x) = 1 - \prod_{i=1}^{k} (1 - g_i(x))^{n_i} \tag{6}$$

Where

$$g_i(x) = \int_{x-\frac{1}{2}l}^{x+\frac{1}{2}l} f_i(t) \, dt \tag{7}$$

and we obtain

$$E(F) = 1 - \frac{1}{L}\int_0^L \prod_{i=1}^{k} (1 - g_i(x))^{n_i} \, dx. \tag{8}$$

The values $g(x)$ are those given in the table and are useful in evaluating the integrals in (5) and (8) by numerical methods.

REFERENCES

[1] LUIS R. SALVOSA, "Tables of Pearson's Type III functions," *Annals of Math Stat*, Vol. 1 (1930), p 191.
[2] NATIONAL BUREAU OF STANDARDS, *Tables of Probability Functions*, Vol. 2 (1942).
[3] H. E ROBBINS, "On the measure of a random set," *Annals of Math. Stat*, Vol. 15, (1944)

CORRECTION TO "A NOTE ON THE FUNDAMENTAL IDENTITY OF SEQUENTIAL ANALYSIS"

BY G E. ALBERT

University of Tennessee

In the paper cited in the title (*Annals of Math. Stat.*, Vol. 18 (1947), pp. 593–596), the proof of Lemma 3 is incorrect. The following correct proof is due to Mr. C. R. Blyth of the Institute of Statistics, University of North Carolina.

It is easy to establish the equation

$$P(n = N|F)[\varphi(t_0)]^{-N} = P(n = N|G)E_{n=N}[\exp(-t_0 Z_N)|G],$$

where $E_{n=N}(u|G)$ denotes the conditional expectation of u under the condition that $n = N$ for any fixed integer N. By Wald [2], equations (2.4) and (2.6), there exists a finite constant C independent of N which dominates the expected values $E_{n=N}[\exp(- t_0 Z_N)|G]$ for every N Thus

$$P(n = N|F)[\varphi(t_0)]^{-N} \leqq C \cdot P(n = N|G). \tag{A}$$

By Stein's theorem [3], there is a positive number t_1 such that $E(\exp nt_1|G)$ is finite But by (A),

$$E\{\exp n[t_1 - \log \varphi(t_0)]\} \leqq C \cdot E(\exp nt_1|G),$$

and Lemma 3 is proved.

CORRECTION TO "ON THE CHARLIER TYPE B SERIES"

BY S. KULLBACK

George Washington University

In the paper cited in the title (*Annals of Math. Stat*, Vol 18 (1947), p. 575), the phrase "so that . . . $R_1 > 1$" on lines 5 and 6 should be deleted I am grateful to Prof. Ralph P Boas, Jr for calling this to my attention.

ABSTRACTS OF PAPERS

Presented June 22–24, 1948 at the Berkeley Meeting of the Institute

1. Estimation of Parameters for Truncated Multinormal Distributions. Z W Birnbaum, E. Paulson and F. C. Andrews, University of Washington.

Let $X_{(N)} = (X_1, \cdots X_p, X_{p+1}, \cdots, X_N)$ be an N-dimensional random variable with a non-singular normal distribution, and let the expectations, variances and covariances of $X_{p+1}, \cdots X_N$ be known. A large sample of $X_{(N)}$ is available, obtained under some side-condition on $(X_{p+1}, \cdots, X_N)$, this side-condition may be a truncation of any kind or, more generally, a selection, i e. imposing on $X_{p+1}, \cdots, X_N$ a probability-distribution different from the original marginal distribution A method is developed for estimating, from such a large sample with a side condition, all the missing parameters of the original distribution of $X_{(N)}$, that is the expectations, variances and covariances of $X_1, \cdots, X_p$, and the covariances $\sigma X_j X_k$ for $j = 1, \cdots, p$ and $k = p+1, \cdots, N$ This method does not require the knowledge of the side-condition (This paper was prepared under the sponsorship of the Office of Naval Research)

2. A Test of the Hypothesis that a Sample of Three Came from the Same Normal Distribution. Carl A. Bennett, General Electric Company, Hanford Works, Richland, Washington.

In the control of the precision of chemical analyses performed in duplicate, a test sometimes becomes necessary as to whether three determinations can reasonably be assumed to have arisen from the same normal population. A critical region for testing this hypothesis is given by $R > R_0$, where $R = D/d$, D being the maximum and d the minimum difference between the three values, and R_0 is determined by integration over the upper tail of the Cauchy distribution. It can easily be seen that this test is equivalent to a t-test between a sample of one and a sample of two

3. A Note on the Application of the Abbreviated Doolittle Solution to Non-Orthogonal Analysis of Variance and Covariance. Carl A. Bennett, General Electric Company, Hanford Works, Richland, Washington

S S Wilks has shown that the sums of squares necessary to the tests commonly made in non-orthogonal analyses of variance or covariance can in general be reduced to the ratio of two determinants If several determinantal operations are performed to remove the singular principal minors, the abbreviated Doolittle solution yields these sums of squares directly A combination of this technique and the calculational methods advanced by Wald and Yates greatly reduces the tedium of calculation in this type of analysis

4. Yield Trials with Backcrossed Derived Lines of Wheat. G A Baker and F. N. Briggs, University of California, Davis.

Strains of White Federation 38 and Baart 38 Wheats derived by backcrossing sufficient to insure a high degree of homogeneity for all genetic factors were grown in conventional yield trials The results were somewhat contradictory and led to a critical examination of such trials The assumption that the deviations of yields in field trials from the specified pattern are random with uniform variance and expectation zero is not sufficiently realistic We are led to consider a mathematical model which assumes a set of fertility levels upon

which a random element is superimposed. On the basis of this model it is possible to account for the low observed correlations between residuals and plot yields. In such a model the variance ratio F may be approximately unbiased but then its variance is smaller than under conventional assumptions. On the other hand, the expected value of F may be greater than one and sufficiently large so that "significant differences" between strains will always be found due to the differences in fertility levels. In such cases the results of the experiment may be misinterpreted. Transformations, in the ordinary sense of the word, will not bring such data into conformity with the conventional model. In order to bring the correlation between residuals and plot yields down to a sufficiently low level it is necessary to concentrate most of the variation in fertility levels into a few plots. That this is not unreasonable is borne out by agronomic observations. This model also explains the absence of correlation between the yields of strains as determined in two different trials on the same set of strains.

5. The Selection of the Largest of a Number of Means. Charles M. Stein, University of California, Berkeley.

Suppose X_{ij}, $i = 1, \cdots, p$; $j = 1, 2, \cdots$ are independently normally distributed with means $\xi_i + \eta_j$ and variances σ_j^2 where ξ_i, η_j are unknown but σ_j^2 are known. ϵ, α are fixed numbers with $0 < \epsilon$, $0 < \alpha < 1$. It is desired to select, by a sequential procedure, in which we take first the observations with second subscript 1, etc. an integer M among $1, \cdots, p$ such that, for every $k = 1, \cdots, p$ and $\xi_1, \cdots \xi_p, \eta_1, \eta_2, \cdots$ satisfying $\xi_k \geqq \xi_j + \epsilon$ for all $j \neq k$, $P\{M = k\} \geqq 1 - \alpha$. In accordance with the following rule, one decides at each stage (after the observations with second subscript n) to take no more observations with certain first subscripts. For each $n = 1, 2, \cdots$ and each $l = 1, \cdots, p$ compute

$$\sum_{j=1}^{n} \frac{1}{\sigma_j^2}\left(X_{l,j} - \bar{X}_j - \frac{\epsilon(t_j - 1)}{t_j}\right)$$

where $\bar{X}_j$ is the average of the observations with second subscript j and t_j is the number of such observations. Continue taking observations $X_{l,n+1}$ for those l for which this expression is greater than $(ln\alpha)/\epsilon$ but not for the others. Eventually there will be at most one subscript $l = 1, \cdots, p$ for which one continues to take observations and if there is one this is chosen to be M. If there is none, the l for which the sum is largest is chosen to be M. This procedure is a straight-forward application of the Lemma on p. 146 of Wald's *Sequential Analysis*, and generalizations can easily be found.

6. The Effect of Inbreeding on Height at Withers in a Herd of Jersey Cattle. W. C. Rollins, S. W. Mead, and W. M. Regan, University of California, Davis.

The data consist of measurements of height at withers of about 200 females for various ages from one month to five years. The intensity of inbreeding as measured by Wright's coefficient of inbreeding averaged 15 per cent and reached as high as 44 per cent in a few cases.

An intra-sire covariance analysis of height and per cent of inbreeding was made for various ages from the first month to the fifty-fourth month.

The results of the statistical analysis indicate that the inbred animals are shorter at one month of age and grow more slowly up to about the sixth month than do the outcrossed animals, but that from the sixth month on the inbreds begin to catch up with the outcrossed so that at maturity there is no significant difference in height.

7. An Example of a Singular Continuous Distribution. HENRY SCHEFFÉ, University of California at Los Angeles.

Simple and "natural" examples of singular continuous probability distributions are of pedagogical interest They are trivially available in the k-variate case for $k > 1$. A univariate example may be obtained from the notion of a sequence of independent trials of an event with constant probability p of success, a notion familiar to the student and indispensable in elementary probability theory. The (real-valued) random variable X is taken to be the dyadic representation of the sequence of results (1 and 0, respectively, for success and failure) It is known that X has a singular continuous distribution for $p \neq 0, \frac{1}{2}, 1$ This result may be proved by using only the Tchebycheff inequality together with the formulas for the mean and variance of the binomial distribution.

8. On the Theory of Some Non-Parametric Hypotheses. ERICH L. LEHMANN and CHARLES STEIN, University of California, Berkeley, California.

For two types of non-parametric hypotheses optimum tests are derived against certain classes of alternatives The two kinds of hypotheses are related and may be illustrated by the following example: (1) The joint distribution of the variables $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ is invariant under all permutations of the variables; (2) the variables are independently and identically distributed. It is shown that the theory of optimum tests for hypotheses of the first kind is the same as that of optimum similar tests for hypotheses of the second kind. Most powerful tests are obtained against arbitrary simple alternatives, and in a number of important cases most stringent tests are derived against certain composite alternatives For the example (1), if the distributions are restricted to probability densities, Pitman's test based on $\bar{y} - \bar{x}$ is most powerful against the alternatives that the X's and Y's are independently normally distributed with common variance, and that $E(X_i) = \xi$, $E(Y_i) = \eta$ where $\eta > \xi$. If $\eta - \xi$ may be positive or negative the test based on $|\bar{y} - \bar{x}|$ is most stringent The definitions are sufficiently general that the theory applies to both continuous and discrete problems, and that tied observations present no difficulties. It is shown that continuous and discrete problems may be combined Pitman's test for example, when applied to certain discrete problems, coincides with Fisher's exact test, and when $m = n$ the test based on $|\bar{y} - \bar{x}|$ is most stringent for hypothesis (1) against a broad class of alternatives which includes both discrete and absolutely continuous distributions

9. Concerning Compound Randomization in the Binary System. JOHN E. WALSH, Project Rand, Santa Monica, California

Consider a set of binary digits. The numerical deviation from $\frac{1}{2}$ of the conditional probability that a specified digit equals 0 is called the bias of that digit for the given conditions on the remaining digits of the set. The maximum bias of the set is defined to be the maximum of the biases of the digits of the set. A set of binary digits is called random if its maximum bias is zero Now consider an array of $(1 + t_1) \cdots (1 + t_K) \times n$ binary digits such that the rows are statistically independent A compounding method of obtaining a set of $t_1 \cdots t_K n$ binary digits from the original array is presented. By suitable choices of $K, t_1, \cdots, t_K$ the maximum bias of the compounded set can be made extremely small even if the maximum bias of the original array is not small; this can be done so that $t_1 \cdots t_K / (1 + t_1) \cdots (1 + t_K)$ is moderately large. Also a method is outlined for constructing an approximately random binary digit table This table has the property that the maximum bias of a set of digits taken from the table is an increasing function of the number of digits in the set.

10. A Multiple Decision Problem Arising in the Analysis of Variance

EDWARD PAULSON, University of Washington, Seattle.

In some applications of the analysis of variance, a procedure is required for classifying varieties into 'superior' and 'inferior' groups. Consider K varieties, with $x_{i\alpha}$ the α^{th} observation on the i^{th} variety ($\alpha = 1, 2, \cdots r, i = 1, 2, \cdots K$), let $\bar{x}_i = \sum_{\alpha=1}^{r} x_{i\alpha}/r$ and let s^2 be an independent estimate of the variance For the i^{th} variety form the corresponding interval $\left(\bar{x}_i - \frac{\lambda s}{\sqrt{r}}, \bar{x}_i + \frac{\lambda s}{\sqrt{r}}\right)$ The superior group then consists of the variety with greatest sample mean, together with those varieties whose corresponding intervals have at least one point in common with the interval for the variety with the greatest mean If all varieties fall into one group, this group is labeled 'neutral' and the varieties are considered homogeneous. To select λ, consider the relative importance of different incorrect classifications For a given λ, an explicit expression is found for $P(A)$, the probability the varieties will not all be classified in one group when $m_1 = m_2 = \cdots = m_K$ where $m_i = E(\bar{x}_i)$, also explicit expressions are found for $P(B_1)$ and $P(B_2)$, where $P(B_1)$ is the probability there will not be a superior group consisting only of the Kth variety and $P(B_2)$ is the probability there will not be a superior group consisting of at least the Kth variety, when $m_1 = m_2 = \cdots = m_{K-1} = m$ and $m_K = m + \Delta (\Delta > 0)$ Similar results are obtained for classifying K processes according to their variances

11. Recurrence Formulae for the Moments and Semi-variants of the Joint Distribution of the Sample Mean and Variance.

OLAV REIERSØL, University of Oslo, Norway.

Let $x_1, x_2, \cdots, x_n$ be independent and having the same distribution We consider the arithmetic mean m and the variance $v = (1/(n-1)) \Sigma (x_i - m)^2$. Let κ_{rs} denote the seminvariants of the joint distribution of m and v, and let the seminvariant generating operators K be defined by the equations: $\kappa_{r+1,s} = K_1 \kappa_{rs}$, $\kappa_{r,s+1} = K_2 \kappa_{rs}$ $K_i' 1 = 0$, $K_i(PQ) = P(K_i Q) + Q(K_i P)$ An operator which operates only on the first factor of a product shall be denoted by a prime, and an operator which operates only on the second factor shall be denoted by a double prime We have the following general formula, valid for any parent distribution $K_1^r[(n-1)(K_2 + \kappa_{01}' + \kappa_{01}'') - 2n(K_1' + \kappa_{10}')(K_1'' + \kappa_{10}'')]^s[1 \cdot (\kappa_{01} - n\kappa_{20}) + n(\kappa_{10}\kappa_{10} - 1 \cdot \kappa_{10}^2)] = 0$. For $s = 0, 1, 2$, we obtain the formulae, $K_1^r(\kappa_{01} - n\kappa_{20}) = 0$, $K_1^r[(n-1)(\kappa_{02} - n\kappa_{21}) - 2n^2\kappa_{20}^2] = 0$, $K_1^r[(n-1)^2(\kappa_{03} - n\kappa_{21}) - 8n^2(n-1)\kappa_{21}\kappa_{20} + 4n^3(n-1)\kappa_{30}^2 - 8n^3(n-1)\kappa_{20}^3] = 0$.

12 The Problem of Identification in Factor Analysis.

OLAV REIERSØL, University of Oslo, Norway.

The paper is concerned with the multiple factor analysis of L. L. Thurstone Thurstone has given criteria which he says are almost certain to constitute sufficient and more than necessary conditions for uniqueness (i e identifiability) of a simple structure It is shown that Thurstone's criteria are not always sufficient, and conditions are derived which are more nearly necessary and sufficient for the identifiability of a simple structure. Let A be the matrix of factor loadings with n rows and r columns. When the communalities are identifiable, the conditions will be. (1) Each column of A should have at least r zeros. (2) Let us consider the submatrix B of A, consisting of all the rows which have zeros in the kth column Then, for $q = 1, 2, \cdots, r - 1$, there should for any combination of q columns different from the kth, exist at least $q + 1$ rows of B containing non-zero elements in the q columns This should be true for any value of k

13 Note on Distinct Hypotheses. AGNES BERGER, Columbia University, New York.

As was pointed out by Neyman, one of the difficulties which one may encounter when devising a test to distinguish between two exhaustive and exclusive composite hypotheses referring to the unknown distribution of a random vector X is the following: If H_0 states that the true distribution function of X belongs to a set $\{F\}$ and H_1 that it belongs to a set $\{G\}$, it may happen that to every Borel set W of the sample space there exists an element F_W in $\{F\}$ and an element G_W in $\{G\}$ for which the probability of the sample point x falling on W is the same and therefore independent of whether H_0 or H_1 is true. If this is the case the pair H_0, H_1 is called non-distinct, otherwise they are called distinct. The existence of non-distinct hypotheses is demonstrated by a simple example, H_0 consisting of one, H_1 of three suitably chosen stepfunctions It is shown however that if the sets $\{F\}$ and $\{G\}$ contain only continuous distribution functions and are at most enumerable then the pair H_0, H_1 is distinct. Necessary and sufficient conditions for H_0 and H_1 to be distinct were obtained jointly with Wald for an important class of hypotheses each containing a continuum of alternatives.

14. Place of Statistical Sampling in the Education of Engineers. E. L GRANT, Stanford University.

There is convincing evidence that many engineering problems could be solved better with the aid of statistical methods than they are now solved without this aid However, few practising engineers or teachers of engineering have had any training in statistical methods. As a result, those engineering problems which are in part statistical problems are seldom recognized as such. Even in the field of industrial quality control, in which successful applications of some of the simpler statistical techniques have been made in many different industries, the surface has barely been scratched and a serious obstacle to progress is the lack of a widespread appreciation of the statistics point of view among design engineers, production engineers, inspection personnel, and management.

This condition might gradually be corrected if during the next few years instruction in statistics should be introduced into all undergraduate engineering curricula. However, some recent discussions touching on the subject of statistics instruction for engineering students (e g , the report on "The Teaching of Statistics" which appeared in the March 1948 issue of the *Annals of Math Stat*) have been most unrealistic regarding the amount of statistics instruction which could be added to engineering curricula These discussions have suggested a full year of basic statistics followed by one or more courses in engineering applications Desirable as this arrangement might be from the point of view of the most effective instruction in statistics, it is out of the question when considered in the light of the many subjects which are needed in engineering curricula Although undergraduate engineering curricula have always been tighter than other curricula, the pressures today are greater than ever before—for more time devoted to the humanistic-social stem, for more time in basic mathematics and science, for introductory courses in various economic and management subjects such as engineering economy, accounting, industrial relations, business law, and industrial organization and management, and for more time in the various departmental courses in engineering subjects in order to permit presentation of important recent developments in engineering technology Under these circumstances the most that can be hoped for in the undergraduate program is a single statistics course for one term, possibly three units for one semester or four units for one quarter. This should be supplemented by additional statistics instruction for some graduate students in engineering A few engineering graduates should be encouraged to take graduate degrees in statistics and to make careers in the field of applied statistics.

In a successful undergraduate statistics course for engineering students, the problems and illustrations should be selected with two purposes in mind. One purpose, of course, should be to develop the principles of probability and statistical method. The other, equally important if these engineering graduates are to persuade their colleagues and superiors to adopt the statistics point of view in approaching engineering problems, should be to demonstrate how statistical method provides a useful guide to action in many different engineering situations. Applications of statistics to industrial quality control provide particularly good problems and illustrative examples which serve this second purpose.

15. Statistical Problems of Medical Diagnosis. JERZY NEYMAN, University of California, Berkeley.

"Diagnosis" is used to describe the outcome of a strictly defined test T, such as Wassermann test, which may lead to either of two possible outcomes, "positive" or "negative". Cases contemplated are such that at the time the test T is performed it is impossible to verify its verdict for certain and the best one can do is to repeat the test. It is postulated that to each individual of a population there corresponds a probability p that the test T will give a positive outcome. The value of p may vary from one individual to another. It is presumed that as p increases, the illness in the patient increases. Problem of comparison of two alternative tests and problem of estimating the distribution of p reduces to problems relating to the distribution of X = number of positive outcomes in n independent diagnoses. Statistical machinery suggested is that of BAN estimates (*Public Health Report*, Vol. 62, (1947), p. 1449). Principal result reported is that, with the mathematical model used in the paper quoted, the empirical variances of four BAN estimates computed for 205 samples of 1000 elements each agreed reasonably with the theoretical asymptotic values. Empirical distributions of three of these estimates did not show deviations from normality. That of the fourth was non-normal. It seems therefore that the asymptotic procedure of BAN estimate may be adequate for similar analyses.

16. Power of Certain Tests Relating to Medical Diagnosis. C. L. CHIANG and J. L. HODGES, JR., University of California, Berkeley.

Associate with each individual in a population π the probability p that he will be found tubercular when examined by a standard X-ray technique. Yerushalmy and others [*J. Am. Med. Assn.*, Vol. 133, (1947), p. 359] performed 5 independent such diagnoses on each of 1256 persons. Neyman [*Public Health Reports*, Vol. 62, (1947), p. 1449] proposed a simple four-parameter model for the distribution of p in π, estimated the parameters from the data of Yerushalmy and others, and obtained a satisfactory fit. In the present paper, the work of Neyman is paralleled with four new models, all giving satisfactory fit with the same data. The five models differ considerably in shape, and in the number of repeated diagnoses which they indicate to be necessary to detect a high proportion of those individuals having, say, $p \geqq 0.1$. Therefore further preliminary study seems indicated before one can design a mass survey to detect a high proportion of such persons. The approximate power of the χ^2 test of the Neyman model is considered, using one of the other models as alternative. It is found that to obtain power 0.7 with level of significance 0.05, it would be necessary to diagnose 5290 individuals 5 times each.

17. Iterative Treatment of Continuous Birth Processes. T. E. HARRIS, Project Rand, Santa Monica, California.

Random variables z_n are defined by $z_0 = 1$, $P(z_1 = r) = p_r$, $r = 1, 2, \cdots$; if $z_n = k$, z_{n+1} is the sum of k independent variates, each distributed like z_1. Let $x = \sum_1^\infty rp_r < \infty$;

$\sum_1^\infty r^2 p_r < \infty$; $0 < p_1 < 1$. The generating function $f(s) = \sum_1^\infty p_r s^r$ is said to be C.I if there exists a family of generating functions $f(s, t)$ with $f(s, 1) = f(s)$, $f[f(s, t), t'] = f(s, tt')$ for all nonnegative t and t'. A necessary and sufficient condition that $f(s)$ be C.I. is that the numbers a_r, $r = 2, 3, \cdots$, be nonnegative, the a_r are determined recursively by requiring that the power series $\xi(s) = -s + \sum_2^\infty a_r s^r$ satisfy formally the functional equation $\xi(s)f'(s) = \xi[f(s)]$. The problem is connected with classical works on iteration. If $f(s)$ is C.I., the given Markoff process can be imbedded in a continuous birth process. If $\xi(s)$ is given, the m.g.f. $\phi(s)$ of the asymptotic distribution of the variate z_n/x^n may be determined from the formula $\phi^{-1}(s) = (s - 1) \exp\left\{\int_1^s \left[\frac{\xi'(1)}{\xi(y)} + \frac{1}{1-y}\right] dy\right\}$. Various properties of the corresponding distribution can be inferred from this expression.

18. Estimation of Means on the Basis of Preliminary Tests of Significance. Blair M. Bennett, University of California, Berkeley.

This paper examines the statistical procedure of pooling two sample means on the basis of the results of one or more preliminary tests of significance. Let x_i, $(i = 1, \cdots, N_1)$, represent a sample of N_1 observations from a normal population $\eta_1(\xi, \sigma_1^2)$, and y_i a sample of N_2 observations from $\eta_2(\eta, \sigma_2^2)$. An estimate of ξ which is commonly used in certain practical situations is given by $x' = \bar{x}$, or $x' = (N_1\bar{x} + N_2\bar{y})/(N_1 + N_2)$, according as the sample means $\bar{x}$, $\bar{y}$ do or do not differ significantly on the basis of a preliminary test. The distribution of the estimate x' is determined, according as $\sigma_1 = \sigma_2$ are known or unknown. In both situations, the maximum (or minimum) bias is computed as a function of various levels of significance of the preliminary test of equality of means. Also, the mean square error of the estimate x' is calculated in both cases. If now equality of variances cannot be assumed, but an F-test of the sample variances s_1^2, s_2^2 does not indicate any significant difference, then in practice $\bar{x}$, $\bar{y}$ may be pooled, the weights being inversely proportional to the sample variances. Thus, the usual estimate of ξ will be of the form: $x' = \bar{x}$, or $x' = (N_1\bar{x}/s_1^2 + N_2\bar{y}/S_2^2)/(N_1/S_1^2 + N_2/S_2^2)$, according as $\bar{x}$ and $\bar{y}$ do or do not differ significantly on the basis of the Student t-test, subsequent to an F-test. The bias and mean square error of this estimate have been computed with the aid of the conditional power function of the t-test subsequent to an F-test.

19. Note on Power of the F Test. Stanley W. Nash, University of California, Berkeley.

Assuming "treatment" expectations to be normal random variables, the ratio of the sum of squares due to treatments to the sum of squares due to error has a central F distribution in the cases of randomized blocks, Latin squares, and one-way classifications. The F statistic converges in probability to a constant as the number of treatments is increased. This is one plus a multiple of the variance between treatment expectations. The power of the F test increases monotonely to one as the number of treatments is increased. This power can be calculated using tables of the incomplete beta function.

20. Best Asymptotically Normal Estimates. E. W. Barankin and J. Gurland, University of California, Berkeley.

The methods of minimum χ^2 developed by Neyman for obtaining BAN (best asymptotically normal) estimates of the parameters appearing in the multinomial distribution

are generalized to obtain certain optimum types of estimates in the case of an arbitrary distribution under certain restrictions. Let the random vector X have the probability density $v(x, \theta)$ in the absolutely continuous case and let $v(x, \theta) = P\{X = x/\theta\}$ in the discrete case, where θ is a fixed vector in the parameter space. Functions $\phi_i(X)$, $(i = 1, 2, \cdots, r)$ are selected for the purpose of forming estimates, these estimates are taken to be functions of the sample moments $\frac{1}{n}\sum_{j=1}^{n} \phi_i(x_j)$. Certain quadratic forms which depend on the choice of functions $\phi_1(X), \phi_2(X), \cdots, \phi_r(X)$ are minimized with respect to the parameters. In this manner, asymptotically normal estimates are obtained which are consistent, and have minimum asymptotic variances within the class of estimates so determined by the particular functions $\phi_1, \phi_2, \cdots, \phi_r$. It is possible, through a modification of this procedure, to obtain estimates by solving a set of linear equations. If $v(x, \theta)$ has the form

$$v(x, \theta) = \exp \{\beta_0(\theta) + \sum_{i=1}^{s} \beta_i(\theta) y_i(x) + y_0(x)\}$$

it can be shown that the best choice of the ϕ's is $y_1(x), y_2(x), \cdots, y_s(x)$. Maximum likelihood estimates belong to this class of BAN estimates.

BOOK REVIEW

The Theory of Games and Economic Behavior *John von Neumann and Oskar Morgenstern.* Princeton University Press, 1947, Second Edition, Pp xviii, 641 $10.00

REVIEWED BY LEONID HURWICZ[1]

Iowa State College

This review is devoted to the second edition of a book which from its first appearance was acknowledged to be a major contribution in the field of theory of rational behavior. As is pointed out in the Preface, "the second edition differs from the first in some minor respects only". The main change is the addition of a proof (of "measurability" of utility) omitted in the first edition.

The book's objective is to solve the problem of rational behavior in a very general type of situation.

It is, therefore, not surprising that its results are of relevance in many fields of knowledge, among them economics and statistical inference.

In both economics and statistics the problem of rational behavior is a fundamental one. Thus one of the classical problems treated by the economic theory is that of profit maximization by a firm. The firm is assumed to be maximizing its net profit which is a function of prices of the product, materials used, etc., as well as the quantities used and produced. In the simplest case prices are taken as given, more generally they are assumed to be functions (known to the firm) of the quantities sold and purchased. But assuming this function to be known presupposes the knowledge of behavior of other firms. This procedure has for a long time been regarded as highly unsatisfactory, it is analogous to elaborating the theory of rational behavior of a poker player on the assumption that he knows the strategy of the other players!

It is the type of situation in which not only the behavior of various individuals, but even their strategies, are interdependent, that is treated by von Neumann and Morgenstern. The essence of their solutions is to base the optimal strategy on the *minimax principle.* As applied to a game, the principle requires that one should choose a strategy which minimizes the maximum loss that could be inflicted by the opponent.

The minimax principle, when applied by both players need not, in general, lead to a stable solution. To ensure the existence of such a solution the authors are led to the postulate that the choice of strategies be made through a random process The minimax to be found is that of the *mathematical expectation* of the loss in the game. The latter postulate is of a restrictive nature[2] since it implies that the game is played for numerical ("measurable") stakes and that

[1] On leave to the United Nations Economic Commission for Europe.

[2] See Jacob C Marschak, "Neumann's and Morgenstern's New Approach to Static Economics", *The Journal of Political Economy,* Vol LIV (1946).

the second and higher moments of the probability distribution of the losses are immaterial. This restriction, however, has permitted the authors to go deeper in other directions. Given the great complexity of the problem, even in its restricted version, the authors' decision can hardly be criticized. One could only wish that similar considerations had made the authors more tolerant towards other work in the field of economics than is shown in some sections of the book.

The readers of the *Annals* will be particularly interested in the connection between the *Theory of Games* and the theory of statistical inference

As has been pointed out by Abraham Wald[3] the problem faced by the statistician is somewhat similar to that of a player in a game of strategy. The theory of statistical inference may be viewed as a theory of rational behavior of the statistician. His "strategy" consists in adopting an optimal test or estimate, more generally an optimal decision function. This optimal decision function must be chosen without the knowledge of the "a priori" distribution of the population parameters. Wald's basic postulate of minimization of maximum risk is equivalent to regarding the statistician as a player in a game of strategy, with "Nature" as the other player. The optimal decision function is chosen in a way which (as shown by Wald) is equivalent to assuming the "least favorable" a priori distribution of the parameters As Wald says, "we cannot say that Nature wants to maximize [the statistician's risk]. However, if the statistician is completely ignorant as to Nature's choice, it is perhaps not unreasonable to base the theory of a proper choice of [the decision function] on the assumption that Nature wants to maximize (the statistician's risk)".

It may be noted, however, that statistical inference, as seen by Wald, is a relatively simple game since it involves only two players and is of the zero-sum variety.

The admiring and enthusiastic reception given to the book's first edition would make any further general appraisal somewhat anticlimatic Suffice it to say that a good deal of valuable work has already been stimulated by the *Theory of Games*, both in the field of social sciences and in mathematics.

[3]Abraham Wald, "Statistical Decision Functions which Minimize the Maximum Risk", *Annals of Mathematics*, Vol 46, (1945).

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Dr. Paul H. Anderson, formerly an Economist with the War Assets Administration, Washington, D. C., has been appointed Professor of Marketing at Loyola University, New Orleans, Louisiana.

Mr. N. H Carrier has resigned his position with the Mathematical Statistics Section, Chief Scientific Advisers Division, Ministry of Works, England to accept an appointment as Statistician in the General Register Office, Somerset House, Strand, London, W. C. 2, England.

Dr. T. Freeman Cope has been promoted to a full professorship at Queens College, Flushing, New York

Dr. Wayne W. Gutzman, who was formerly at the Postgraduate School, Naval Academy, Annapolis as an Assistant Professor, has accepted a professorship in the Department of Mathematics, University of South Dakota.

Mr. Elvin A Hoy has transferred from the position as Chief, Statistics Section, Bureau of Research and Statistics in the Social Security Administration to the position as Chief, Research Evaluation Section, Naval Reserve Training Publications, Navy Department, Naval Gun Factory, Washington, D. C.

Dr Joe J. Livers has been promoted to a full professorship at Montana State College, Bozeman, Montana.

Professor Ernest S Keeping has returned to his position at the University of Alberta, Edmonton, Alberta, Canada after having spent the spring term of 1948 at the Institute of North Carolina

Mr. Wharton F. Keppler of the M&R Dietetic Laboratories, Inc , Columbus, Ohio has recently qualified as a Professional Industrial Engineer in the State of Ohio.

Mr. Ralph Mansfield has formed his own company to manufacture electrical testing equipment. The company is known as the Auto-Test, Incorporated, with Mr Mansfield acting as Vice-president and Chief Engineer.

Mr Jack Moshman has resigned an instructorship in mathematics at the University of Tennessee to accept the post of Statistician to the Medical Advisor, United States Atomic Energy Commission, Oak Ridge, Tennessee.

Mr. Bernard E. Phillips has resigned his position as teacher of mathematics in the Newark, New Jersey high schools to do statistical work for the Glenn L. Martin Co , Baltimore, Maryland

Dr. W. R. Van Voorhis, Associate Professor of Mathematics, Fenn College, attended, as a representative of the Institute of Mathematical Statistics, the inauguration ceremonies of Dr Keith Glennan as President of Case Institute of Technology, Cleveland, Ohio.

Atomic Energy Commission Fellowships

The National Research Council is announcing a new program of fellowships supported by funds provided by the Atomic Energy Commission as a part of the Commission's responsibility for future atomic energy research. Accordingly, these fellowships will be awarded in the many fields of science related to research in atomic energy

A considerable number of these fellowships is available to young men and women who wish to continue in graduate training or research for the doctorate in an appropriate field of science. Others of these fellowships will provide training in biophysics applied to the control of radiation hazards. An additional number of fellowships will be assigned to those below the age of 35 who have already achieved the doctorate and who wish to secure advanced research training and experience in those aspects of the physical, biological and medical sciences related to atomic energy Tenure of the fellowship does not impose on the fellow any commitment with regard to subsequent employment.

The candidates will be selected by the fellowship boards of the National Research Council established for this program In the postdoctoral field, there will be three groups of fellowships, the basic stipend of which will be $3000. For the selection of fellows for advanced research and training in the general field of the physical sciences, a board has been established with Dr. Roger Adams, Professor of Chemistry, University of Illinois, as chairman In the general field of the biological sciences, exclusive of the medical sciences, selections of postdoctoral fellows will be made by a board under the chairmanship of Dr. R. G. Gustavson, Chancellor of the University of Nebraska. For the selection of postdoctoral fellows in the medical sciences, a board has been set up under the chairmanship of Dr Homer W. Smith, Professor of Physiology, College of Medicine, New York University.

The program provides for two groups of fellows in the predoctoral field, with stipends ranging from $1500–2100 One group of fellows will work in the biological and basic medical sciences including applied biophysics related to the measurement and control of radiation hazards and the effect of radiation upon health. Selections will be made by a board under the chairmanship of Dr. Douglas Whitaker, Professor of Biology, and Dean of the School of Biological Sciences, Stanford University. Another group of predoctoral fellows will be selected for study and research in the general field of the physical sciences. Selections will be made by a board under the chairmanship of Dr. Henry A. Barton, Director of the American Institute of Physics.

Fellowships will be granted for study and research in universities or other nonprofit research establishments approved by the fellowship boards Awards will be made for the academic year 1948–49. Supervision of a fellow's program of work will be under the direction of the fellowship boards of the National Research Council. Further information can be secured by writing to the Fellowship Office, National Research Council, 2101 Constitution Avenue, Washington 25, D. C.

Research Fellowships in Psychometrics

The Educational Testing Service, Princeton, N. J., is offering for 1949–50 its second series of research fellowships in psychometrics leading to the Ph.D degree at Princeton University Open to men who are acceptable to the Graduate School of the University, the two fellowships carry a stipend of $2,200 a year and are normally renewable

Fellows will be engaged in part-time research in the general area of psychological measurement at the offices of the Educational Testing Service and will, in addition, carry a normal program of studies in the Graduate School. Competence in mathematics and psychology is a prerequisite for obtaining these fellowships. Information and application blanks may be obtained from: Director of Psychometric Fellowship Program, Educational Testing Service, Box 592, Princeton, N. J.

Preliminary Actuarial Examinations

Prize Awards

The winners of the prize awards offered by the Actuarial Society of America and the American Institute of Actuaries to the nine undergraduates ranking highest on the combined score on Part 1 and Part 2 of the 1948 Preliminary Actuarial Examinations are as follows·

First Prize of $200

Edward H Larson	*Massachusetts Institute of Technology*

Additional Prizes of $100

John E Brownlee	*Haverford College*
William L Farmer	*University of Alabama*
Joseph P. Fennell	*Princeton University*
Bert F Green, Jr	*Yale University*
Solomon Leader	*Rutgers University*
Felix A. E Pirani	*University of Western Ontario*
Richard J Semple	*University of Toronto*
Charles A Yardley	*Dartmouth College*

The two actuarial organizations have authorized a similar set of nine prize awards for the 1949 Examinations on Part 2.

The Preliminary Actuarial Examinations consist of the following three examinations:

Part 1. *Language Aptitude Examination.*

(Reading comprehension, meaning of words and word relationships, antonyms, and verbal reasoning.)

Part 2 *General Mathematics Examination.*

(Algebra, trigonometry, coordinate geometry, differential and integral calculus.)

Part 3 *Special Mathematics Examination*

(Finite differences, probability and statistics.)

The 1949 Preliminary actuarial Examinations will be administered by the Educational Testing Service at centers throughout the United States and Canada on May 13–14, 1949. The closing date for applications is March 15, 1949

Detailed information concerning the Examinations can be obtained from either of the following organizations:

American Institute of Actuaries,
135 South LaSalle Street,
Chicago 3, Illinois.

The Actuarial Society of America,
393 Seventh Avenue,
New York 1, New York.

New Members

The following persons have been elected to membership in the Institute

(March 1 to May 31, 1948)

Alder, Arthur, Ph.D (Univ. of Berne) Professor of Actuarial Science, University of Berne, Schlaeflistrasse 2, Berne, Switzerland.

Andrews, Fred C., B S. (Univ of Washington) Research Fellow, Department of Mathematics, University of Washington, *141 Savery Hall, University of Washington, Seattle, Washington.*

Archer, John, Actuary, Pensions Section, Lever Brothers and Unilever Ltd , *5A Spencer Hill, Wimbledon, S. W 19, England*

Benitz, Paul A., M.A (Stanford Univ) *173 Serpentine Road, Tenafly, New Jersey.*

Bennett, George K., Ph D (Yale) President of the Psychological Corporation, 522 Fifth Avenue, New York 18, New York.

Berrettoni, J. N., Ph.D (Univ of Minnesota) Professional Consultant in Statistics and Economics, *632 Erie St , S. E , Minneapolis 14, Minnesota.*

Birnbaum, Allan, A.B. (Univ. of Calif., Los Angeles) Teaching Assistant, Mathematical Statistics Department, Columbia University, *500 Riverside Drive, Room 434, New York 27, New York*

Blank, Mark, M A. (Univ of Pennsylvania) Instructor of Philosophy, University of Pennsylvania, *223 E Sedgwick, Philadelphia, Pa*

Blumen, Isadore, B A. (Univ. of Minn) Student, Department of Mathematical Statistics, University of North Carolina, Chapel Hill, North Carolina

Burdick, Wayne E., M A. (Univ of Mich) Student, University of Michigan, *314 S Fifth Avenue, Ann Arbor, Michigan*

Chaturvedi, Jagdish C., M.Sc. (Agra Univ , India) Lecturer in Statistics, St John's College, *37, Delhi Gate, Agra, U P , India.*

Cote, Louis J., A.M (Univ of Mich) Student, University of Michigan, *315 North State Street, Ann Arbor, Michigan.*

Dunleavy, Mary, A B (Hunter College, New York) Statistician, E I. Dupont de Nemours, *657 Second Avenue, New York 16, New York*

Ferber, Robert, M.A (Univ of Chicago) Student, University of Chicago, 54 West 89th Street, New York 24, New York

Forman, John W., M S (Univ of Iowa) Graduate Assistant, Department of Mathematics, State University of Iowa, Iowa City, Iowa.

Franklin, Nathan M., M.S. (Univ. of Mich.) Student, Univ. of Michigan, Box 195, *Moodus, Connecticut.*

Fraser, Donald A. S., M A. (Univ. of Toronto) Instructor in Statistics, Graduate College, Princeton, New Jersey

Grabowski, Edwin F., A.B. (George Washington Univ.) Student, George Washington University, *1330-30th Street, N.W., Washington, D. C.*

Healy, William C., Jr., B.S.E. (Univ of Mich.) Student, University of Michigan, *589 Lincoln, Grosse Pointe, Michigan.*

Heimdahl, Olaf E. W., A.B. (Luther College, Washington) Teaching Fellow, Department of Mathematics, University of Washington, *4236 Union Bay Lane, Seattle 5, Washington.*

Henriksen, Robert O., B.Sc. (Univ. of Mich) Student, University of Michigan, *751 Clancy Avenue, Grand Rapids, Michigan.*

Howard, William G., B.S. (Western Carolina Teachers College, Cullowhee, N C.) Student, Institute of Statistics, University of North Carolina, *Route 1, Morrisville, North Carolina.*

Irick, Paul E., M S. (Purdue Univ.) Mathematics Instructor, Purdue University, *729 North Grant St., West Lafayette, Indiana*

Johnson, Elgy S., M.A (Univ. of Mich.) Student, University of Michigan, *13907 Lincoln Street, Detroit 3, Michigan.*

Kaplan, E. L., B S. (Carnegie Inst of Tech.) Mathematician, Naval Ordnance Laboratory, *1427 N. St., N. W., Washington 5, D. C.*

Kaufman, Arthur, M.A. (Columbia Univ.) Student and Lecturer of Mathematics, Columbia University, *1280 Sheridan Avenue, New York 56, New York.*

Link, Richard F., B.S. (Univ of Oregon) *750 W Sixth St., Eugene, Oregon.*

Marks, Charles L., M.A (Univ of North Carolina) Instructor of Mathematics, University of North Carolina, *213 Mangum Dormitory, University of North Carolina, Chapel Hill, North Carolina*

Marquardt, Mary, M.A. (Univ. of Illinois) Assistant Professor of Statistics, New York State School of Industrial and Labor Relations, Cornell University, Ithaca, New York.

Mickey, Max R., Jr., B.S (Virginia Polytechnic Institute) Graduate Student and Graduate Assistant, Department of Mathematics, Iowa State College, *706 Ash Avenue, Ames, Iowa.*

Mindlin, Albert, B A. (Univ. of California, Los Angeles) Teaching Assistant, Mathematics Department, University of California, *2444 Carlston Street, Berkeley 4, California.*

Morris, William S., A B (Princeton) Statistician, First Boston Corporation, 100 Broadway, New York 5, New York

Netzorg, Merton J., Engineer, Development Tire Engineering Department, U S Rubber Co , Detroit, Michigan, *2523 Gladstone, Detroit 6, Michigan*

Norton, James A., Jr., A.B. (Antioch College) Graduate Research Assistant, Veterans Guidance Center, Purdue University, West Lafayette, Indiana.

Perrin, John K., A B. (Columbia College) Assistant Statistician, American Telephone & Telegraph Co., 195 Broadway, New York 7, New York.

Perry, Norman C., M.A (Univ of Southern Calif.) Lecturer in Mathematics, Mathematics Department, University of Southern California, Los Angeles, California.

Powell, Charles Jr., Actuary, Coates and Herfurth, Consulting Actuaries, 116 S Virgil Avenue, Los Angeles 4, California.

Raiffa, Howard, B.S. (Univ. of Mich) Student, University of Michigan, *1447 Enfield Court, Willow Run Village, Michigan.*

Raup, Joan E., B A (Barnard College) Statistical Analyst, Bureau of the Budget, *1436 N Street, N W., Washington 5, D. C.*

Rubinstein, David, B S. (Univ. of Wash) Research Assistant, Statistical Laboratory, University of California, *2216 Parker Street, Berkeley 4, California*

Schlenz, John W., B S. (Baldwin-Wallace College) Student, University of Michigan, *8306 Vineyard Avenue, Cleveland 5, Ohio.*

Scott, Elizabeth L., A B (Univ of California) Research Assistant, Statistical Laboratory, Department of Mathematics, University of California, Berkeley 4, California

Seidman, Herbert, B.A. (Brooklyn College) Junior Statistician, Chief, Statistical Information Section, New York University and Student, New York University, *2170 New York Avenue, Brooklyn 10, New York.*

Shaw, Oliver A., B A. (Univ. of Mississippi) U.S Air Force, *6431 Brooks Lane, N W., Washington, D. C*

Shellard, Gordon D., B.S. (Mass Institute of Tech) Assistant Section Head, Underwriting Studies Section, Actuarial Division, Metropolitan Life Insurance Co., *420 Mountain Avenue, Ridgewood, New Jersey.*

Shepherd, Clarence M., M.S (Case Institute of Tech) Electrochemical Research Chemist, *3959 Nichols Avenue, S. W , Washington, D. C*

Shrikhande, Sharad-Chandra S., B Sc. (Nagpur Univ , India) Graduate student, Department of Mathematical Statistics, University of North Carolina, Chapel Hill, North Carolina

Sirlin, Robert, M.A. (Columbia Univ.) Statistician, Financial Analysis, *2046 East 23rd Street, Brooklyn 29, New York.*

Stacy, Edney W., A B. (Univ. of North Carolina) Instructor of Mathematics, University of North Carolina, *301 W. Franklin Street, Chapel Hill, North Carolina*

Sternhell, Charles M., B.S (College of City of N. Y.) Section Head, Actuarial Division, Metropolitan Life Insurance Co , 1 Madison Avenue, New York City, New York

Tang, Pei-Ching, Ph D. (Univ. College, London Univ.) Professor, National Central University, Nanking, China

Whitson, Milo E., A.M. (Geo Peabody College, Nashville) Head of Mathematics Department, California State Polytechnic College, *523 Lawrence Dr , San Luis Obispo, California*

Watson, Geoffrey S., B.A. (Univ. of Melbourne) Student, Institute of Statistics, State College, Raleigh, North Carolina

Woolsey, Theodore D., B.A (Yale Univ.) Biostatistician, Division of Public Health Methods, U S Public Health Service, *111 West Underwood St , Chevy Chase 15, Maryland.*

Wymer, John P., M A (Univ of California, Berkeley) Statistician, U S Bureau of Labor Statistics, *719 Whittier St , N W , Washington 12, D C*

Yerushalmy, Jacob, Ph.D. (Johns Hopkins Univ.) Professor of Biostatistics, School of Public Health, University of California, Berkeley 4, California.

REPORT ON THE BERKELEY MEETING OF THE INSTITUTE

The Thirty-fourth Meeting and the Third Regional West Coast Meeting of the Institute of Mathematical Statistics was held on the Berkeley Campus of the University of California June 22nd through June 24th, 1948, in conjunction with the Twenty-ninth Annual Meeting of the Pacific Division of the American Association for the Advancement of Science. During the meeting 115 persons registered, including the following members of the Institute:

G. A. Baker, Blair M. Bennett, Carl A. Bennett, Z. Wm. Birnbaum, David Blackwell, Albert H. Bowker, George W. Brown, A. George Carlton, Douglas G. Chapman, Andrew G. Clark, Edwin L. Crow, Dorothy Cruden, Harold Davis, R. C. Davis, W. J. Dixon, Robert Dorfman, George Eldredge, Lillian Elveback, Mary Elveback, Benjamin Epstein, M. W. Eudey, Evelyn Fix, Merrill M. Flood, H. H. Germond, Meyer A. Girshick, Eugene L. Grant, John Gurland, T. E. Harris, J. L. Hodges, Jr., Paul G. Hoel, John M. Howell, Harry M. Hughes, Leo Katz, H. S. Konijn, T. C. Koopmans, George W. Kuznets, E. L. Lehmann, Richard F. Link, A. M. Mood, Stanley W. Nash, J. Neyman, Stefan Peters, G. Baley Price, Kathryn B. Rolfe, Leonard J. Savage, Henry Scheffé, Howard L. Schug, Elizabeth L. Scott, Esther Seiden, Milton Sobel, Zenon Szatrowski, John W. Tukey, J. R. Vatnsdal, A. Wald, John E. Walsh, Holbrook Working, Zivia S. Wurtele.

The Tuesday morning session was devoted to a symposium on *Mathematical Theory of Games* with Professor G. C. Evans of the University of California, Berkeley, as chairman. Addresses were:

1. *Survey of von Neumann's mathematical theory of games.* J. C. C. McKinsey, Project Rand.
2. *Recent developments in the mathematical theory of games.* Olaf Helmer, Project Rand.
3. *Applications of theory of games to statistics.* Abraham Wald, Columbia University.
4. *On continuous games.* Henri F. Bohnenblust, California Institute of Technology.
5. *Discussion.* Edward W. Barankin, University of California, Berkeley.

The attendance was approximately 130.

The Tuesday afternoon session was under the chairmanship of Professor Henri F. Bohnenblust of the California Institute of Technology. The invited address, *Complete Classes of Statistical Decision Functions,* by Professor Abraham Wald was followed by tea in Senior Women's Hall and then the following contributed papers:

1. *Identification as a problem of inference.* T. C. Koopmans, Cowles Commission for Research in Economics.
   Discussion: Olav Reiersøl, University of Oslo.
2. *Estimation of parameters for truncated multinormal distributions.* Z. W. Birnbaum, E. Paulson and F. C. Andrews, University of Washington.
3. *A test of the hypothesis that a sample of three came from the same normal distribution.* Carl A. Bennett, General Electric Company.
4. *A Note on the application of the abbreviated Doolittle solution to nonorthogonal analysis of variance and covariance.* (By title.) Carl A. Bennett, General Electric Company.

The attendance was between 100 and 150 during the afternoon.

The Wednesday morning session was devoted to a symposium on *Design of Experiments with Particular Reference to Agricultural Trials* Dean A. R Davis of the University of California, Berkeley, presided briefly and then Professor Abraham Wald took over the duties of chairman The papers were·

1 *Recent advances in experimental design.* R C Bose, University of Calcutta.
2 *Yield trials with backcrossed derived lines of wheat* G A Baker and F N Briggs, University of California at Davis
3. *Selecting subset which includes the largest of a number of means* Charles Stein, University of California, Berkeley
4. *Discussion* A. G Clark, Colorado State College, S W. Nash, University of California, Berkeley, J R Vatnsdal, State College of Washington
5 *The effect of inbreeding on height at withers in a herd of Jersey cattle* W C. Rollins, S W Mead and W. M Regan, University of California at Davis

Attendance was about 100.

The afternoon session, under the chairmanship of Professor George Pólya of Stanford University, began with an invited address by Professor Michel Loève, University of California, Berkeley, on *Random Functions and Related Problems* This was followed by the contributed papers:

1. *An example of a singular continuous distribution* (By title) Henry Scheffé, University of California at Los Angeles.
2 *On the theory of some nonparametric hypotheses* E. L Lehmann and Charles Stein, University of California, Berkeley.
3. *Compound randomization in the binary system* John E. Walsh, Project Rand
4 *A multiple decision problem arising in the analysis of variance* Edward Paulson, University of Washington.
5 *Recurrence formulae for the moments and seminvariants of the joint distribution of the sample mean and variance* Olav Reiersøl, University of Oslo
6. *Identification problem in factor analysis* (By title) Olav Reiersøl, University of Oslo
7. *On distinct hypotheses.* Mrs. Agnes Berger, Columbia University

The attendance was approximately 100.

A symposium on *Sampling for Industrial Use* occupied the Thursday morning session. Professor Z. W Birnbaum of the University of Washington presided.

1 *Sampling plans for continuous production.* M A Girshick, Project Rand.
2 *Sampling plans with continuous variables for acceptance inspection* A L Bowker, Stanford University
3. *Place of statistical sampling in the education of engineers* E L Grant, Stanford University
4. *Discussion.* Henry Scheffé, University of California at Los Angeles; Charles Stein, University of California, Berkeley; Holbrook Working, Stanford University.

The attendance was approximately 100.

The first part of the afternoon session, presided over by Professor W. J. Dixon, University of Oregon, was devoted to contributed papers:

1 *Statistical problems of medical diagnosis.* Jerzy Neyman, University of California, Berkeley.
Discussion. L. J. Savage, University of Chicago

2. *Power of certain tests relating to medical diagnosis.* C. L. Chiang and J. L. Hodges, University of California, Berkeley.
3. *On best asymptotically normal estimates* Edward W Barankin and John Gurland, University of California, Berkeley.
4. *Iterative treatment of continuous birth processes.* T. E Harris, Project Rand
5. *Estimation of means on the basis of preliminary tests of significance* Blair M Bennett, University of California, Berkeley. (By title.)

The attendance was about 90

The second part of the afternoon session was the Business Meeting. Professor Abraham Wald, President of the Institute, presided. It was recommended that two meetings a year be held on the West Coast, one in June in the San Francisco Bay Area alternating between Berkeley and Stanford and the other during the winter alternating between the North and Los Angeles. The next West Coast meeting will be held during the Thanksgiving recess at Seattle.

THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. XIX, No. 4 — December, 1948

THE ANNALS OF MATHEMATICAL STATISTICS

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics, Mt. Royal & Guilford Aves., Baltimore 2, Md. Subscriptions, renewals, orders for back numbers and other business communications should be sent to the ANNALS OF MATHEMATICAL STATISTICS, Mt. Royal & Guilford Aves, Baltimore 2, Md., or to the Secretary of the Institute of Mathematical Statistics, P. S. Dwyer, 116 Rackham Hall, University of Michigan, Ann Arbor, Mich.

Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue. The months of issue are March, June, September and December.

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $8.00 inside the Western Hemisphere and $5.00 elsewhere. Single copies $3.00. Back numbers are available at $8.00 per volume or $3.00 per single issue.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879

TESTING COMPOUND SYMMETRY IN A NORMAL MULTIVARIATE DISTRIBUTION

By David F. Votaw, Jr.

Yale University

Summary. In this paper test criteria are developed for testing hypotheses of "compound symmetry" in a normal multivariate population of t variates ($t \geq 3$) on basis of samples. A feature common to the twelve hypotheses considered is that the set of t variates is partitioned into mutually exclusive subsets of variates. In regard to the partitioning, the twelve hypotheses can be divided into two contrasting but very similar types, and the six in one type can be paired off in a natural way with the six in the other type. Three of the hypotheses within a given type are associated with the case of a single sample and moreover are simple modifications of one another, the remaining three are direct extensions of the first three, respectively, to the case of k samples ($k \geq 2$) The gist of any of the hypotheses is indicated in the following statement of one, denoted by $H_1(mvc)$: *within each subset of variates the means are equal, the variances are equal and the covariances are equal and between any two distinct subsets the covariances are equal.*

The twelve sample criteria for testing the hypotheses are developed by the Neyman-Pearson likelihood-ratio method. The following results are obtained for each criterion (assuming that the respective null hypotheses are true) for any admissible partition of the t variates into subsets and for any sample size, N, for which the criterion's distribution exists: (i) the exact moments, (ii) an identification of the exact distribution as the distribution of a product of independent beta variates, (iii) the approximate distribution for large N. Exact distributions of the single-sample criteria are given explicitly for special values of t and special partitionings.

Certain psychometric and medical research problems in which hypotheses of compound symmetry are relevant are discussed in section 1 Sections 2–6 give statements of the hypotheses and an illustration, for $H_1(mvc)$, of the technique of obtaining the moments and identifying the distributions. Results for the other criteria are given in sections 7–8. Illustrative examples showing applications of the results are given in section 9.

1. Introduction. In studying psychological examinations, or other measuring devices, one may wish to test whether several forms of an examination may be used interchangeably. Consider the case of three forms, and assume that scores of individuals on the three forms have a normal 3-variate distribution The hypothesis of interchangeability is equivalent to the hypothesis that in the normal distribution the means are equal, the variances are equal, and the covariances are equal. When this hypothesis is true, the normal distribution is in-

variant over all permutations of the variates and is said to have *complete symmetry*. It is frequently more important, however, not only to test that the forms have completely symmetric relations with each other but also that they are interchangeable with regard to their relation to some outside criterion measure (e g., the criterion might be skill in a given task). Assuming that the scores of individuals on the three forms and the criterion have a normal 4-variate distribution, the hypothesis of interchangeability is equivalent to the hypothesis of equality of means, equality of variances, and equality of covariances among the three forms and equality of covariances between forms and criterion. When this hypothesis is true, the 4-variate normal distribution is invariant over all permutations of the three variates (associated with forms) among themselves, and so the variance-covariance matrix has the following form:

$$\left\| \begin{array}{c:ccc} A & C & C & C \\ \hdashline C & B & D & D \\ C & D & B & D \\ C & D & D & B \end{array} \right\|,$$

where the quantity A represents the variance of the criterion measure. A normal distribution for which this hypothesis is true is said to have *compound symmetry* (of type I). A more general case of compound symmetry (of type I) arises when there are several examinations (no two of which need have the same number of forms) and several outside criteria

The hypothesis of complete symmetry may arise in certain medical-research problems. For example, suppose a measurement (e.g., $\%CO_2$ in blood) is made at each of three times (say T_1, T_2, T_3) on an experimental animal and assume that the three quantities have a normal 3-variate distribution; one may then be interested in testing the hypothesis of complete symmetry on basis of measurements (considered as a random sample) made on several experimental animals More generally, let there be two characteristics, say U and W (e.g., $\%CO_2$ in blood and $\%O_2$ in blood), which are both measured at each of two times, T_1, T_2. Let it be assumed that the four quantities—which we represent as UT_1, UT_2, WT_1, WT_2—have a normal 4-variate distribution. One may then be interested in testing the hypothesis that the means of the first two variates are equal, the means of the second two are equal, and the variance-covariance matrix has the form:

$$\left\| \begin{array}{cc:cc} E & F & K & L \\ F & E & L & K \\ \hdashline K & L & G & J \\ L & K & J & G \end{array} \right\|.$$

When this hypothesis is true, the 4-variate distribution is said to have *compound symmetry* (of type II). A more general case of compound symmetry (of type II) arises when there are h characteristics and n times ($h, n = 2, 3, \cdots$).

Either of the two types of compound symmetry is a direct extension of complete symmetry. Wilks [5] has thoroughly treated the sampling theory of certain criteria for testing various hypotheses of complete symmetry regarding a normal multivariate distribution.

The problems dealt with in this paper are: (i) to give sample criteria for testing hypotheses of compound symmetry regarding a normal multivariate distribution, and (ii) to give the moments and identify the distribution of each sample criterion when the corresponding hypothesis is true

The hypotheses are stated in section 2 Certain properties of compound symmetric normal distributions are given in section 3 Sections 4, 5, and 6 together give the method of deriving each sample criterion and the methods of obtaining the criterion's moments and identifying its distribution; the methods are illustrated for one of the hypotheses Sections 7–8 give the other criteria and their moments together with approximate distributions of the criteria for large sample sizes Exact distributions of some of the criteria are given in section 7g for certain special compound symmetries Section 9 contains two illustrative examples

2. Statements of hypotheses. Let Π be a normal t-variate population and X_i $(i = 1, \cdots, t)$ $(t \geq 3)$ be the i-th variate in Π. Let the set of variates X_1, $X_2, \cdots, X_t$ be partitioned into q mutually exclusive subsets of which, say, b subsets contain exactly one variate each and the remaining $q - b = h$ subsets (where $h \geq 1$) contain $n_1, n_2, \cdots, n_h$, variates, respectively, where $n_a \geq 2$ $(a = 1, \cdots, h; b + \sum_{a=1}^{h} n_a = t)$. No generality is lost in assuming that the t variates are ordered so that the first b belong to the b subsets containing one variate each, the next n_1 variates belong to the $(b + 1)$-th subset, $\cdots$, the last n_h variates to the q-th subset, where $n_1 \leq n_2 \leq \cdots \leq n_h$. Let $(1^b, n_1, n_2, \cdots, n_h)$ represent such a partition of the variates $X_1, \cdots, X_t$ into subsets; when $b = 0$ the term 1^b will be omitted. The notation can be simplified when $n_1, n_2, \cdots, n_h$ are not all unequal, e g., $(1^b, 2, 2)$ can be written as $(1^b, 2^2)$.

In the statement of each of the following six hypotheses it is assumed that there is a preassigned partition $(1^b, n_1, n_2, \cdots, n_h)$ of the t variates into q subsets $(q = b + h)$.

(1) $H_1(mvc)$: The hypothesis that within each subset of variates the means are equal, the variances are equal, and the covariances are equal and that between any two distinct subsets of variates the covariances are equal.

(2) $H_1(vc)$: The hypothesis that within each subset of variates the variances are equal and the covariances are equal and that between any two distinct subsets of variates the covariances are equal.

(3) $H_1(m)$: The hypothesis that within each subset of variates the means are equal, given that $H_1(vc)$ is true.

(4) $H_k(MVC \mid mvc)$: the hypothesis that k normal t-variate distributions are the same given that they all satisfy $H_1(mvc)$ for a particular division of the variates into subsets $(k \geq 2)$.

(5) $H_k(VC \mid mvc)$: The hypothesis that k normal t-variate distributions have the same variance-covariance matrix, given that they all satisfy $H_1(mvc)$ for a particular division of the variates into subsets ($k \geq 2$).

(6) $H_k(M \mid mVC)$: The hypothesis that k normal t-variate distributions are the same, given that they all satisfy $H_1(mvc)$ for a particular division of the variates into subsets and that they all have the same variance-covariance matrix ($k \geq 2$).

Any of the hypotheses stated above can be expressed in terms of an invariance condition on the normal t-variate distribution (or distributions), e.g., $H_1(mvc)$ is equivalent to the hypothesis that the distribution is invariant over all permutations of the variates within subsets. The pattern of symmetry present in the variance-covariance matrix of the distribution when any of the above six hypotheses is true is given in section 3 (see (3.2)).

Six additional hypotheses, $\bar{H}_1(mvc)$, $\bar{H}_1(vc)$, $\cdots$, $\bar{H}_k(M \mid mVC)$, which are modifications of $H_1(mvc)$, $H_1(vc)$, $\cdots$, $H_k(M \mid mVC)$, respectively, will also be considered. In regard to any of these six $\bar{H}$ hypotheses, it is assumed that there is a partition $(n^h)(n = 2, 3, \cdots)$ of the t variates ($t = nh$) and that in each subset the variates are in a given order; thus each subset has n variates and between any two distinct subsets of variates there are n^2 covariances, which form an $n \times n$ "block" in the variance-covariance matrix of the distribution (see (3.4)). The hypotheses may now be stated as follows:

$\bar{H}_1(mvc)$: The hypothesis that within each subset of variates the means are equal, the variances are equal, and the covariances are equal and that between any two distinct subsets of variates the diagonal covariances are equal and the off-diagonal covariances are equal.

$\bar{H}_1(vc)$: The hypothesis that within each subset of variates the variances are equal and the covariances are equal and that between any two distinct subsets of variates the diagonal covariances are equal and the off-diagonal covariances are equal.

The statement of any of the hypotheses $\bar{H}_1(m)$, $\bar{H}_k(MVC \mid mvc)$, $\bar{H}_k(VC \mid mvc)$, and $\bar{H}_k(M \mid mVC)$ is obtained from the statement of the corresponding H hypothesis by simply substituting $\bar{H}$ for H. The pattern of symmetry present in the variance-covariance matrix of the distribution when any of the six $\bar{H}$ hypotheses is true is given in section 3 (see (3.4)), from which the appropriate invariance condition on the normal distribution can be obtained.

A test of any of the hypotheses $H_1(mvc)$, $\bar{H}_1(mvc)$, $H_1(vc)$, $\bar{H}_1(vc)$, $H_1(m)$, $\bar{H}_1(m)$ is based on a random sample from a normal multivariate distribution; a test of any of the remaining hypotheses is based on k random samples from k normal multivariate distributions, respectively, ($k \geq 2$)

A normal distribution for which an H or $\bar{H}$ hypothesis is true will be called *compound symmetric*. In the special case where the compound symmetry holds for a partition (t) of the t variates, any H hypothesis and the $\bar{H}$ hypothesis corresponding to it are identical; in this case the normal distribution will be called *completely symmetric*. Problems (i) and (ii) (see section 1) have been

solved completely by Wilks [5] for $H_1(mvc)$, $H_1(vc)$, and $H_1(m)$ for the case of complete symmetry.

3. Block symmetric matrices and vectors. Let m_i be the mean value of X_i and $\|\rho_{ij}\sigma_i\sigma_j\|$ be the variance-covariance matrix of $X_1, \cdots, X_t$ $(i, j = 1, \cdots, t)$ (ρ_{ij} is the coefficient of correlation between X_i and X_j). The joint probability density function[1] of $X_1, X_2, \cdots, X_t$ is

$$f(X_1, X_2, \cdots, X_t) = |G_{ij}|^{1/2}\pi^{-t/2}\exp\left[-\sum_{i,j} G_{ij}(X_i - m_i)(X_j - m_j)\right], \tag{3.1}$$

where $\|G_{ij}\|$ is positive definite and its inverse $\|G^{ij}\| = \|2\rho_{ij}\sigma_i\sigma_j\|$.

When any of the H hypotheses is true (see section 2), we represent $\|G^{ij}\|$ by $\|A^{ij}\|$ (also $\|G_{ij}\|$ by $\|A_{ij}\|$) which can be written as (3.2) (see page 452), where $A^{ss'} = A^{s's}$ $(s, s' = 1, \cdots, b)$ and $D^{aa'} = D^{a'a}$ $(a, a' = 1, \cdots, h, a \neq a')$. The A's and B's with single superscripts and the C's and D's have been introduced to indicate the *block pattern* clearly. In general $C^{sa} = C^{as}$ only if $a = s$ $(s = 1, \cdots, b;\ a = 1, \cdots, h)$. $\|A_{ij}\|$ and $\|A^{ij}\|$ have the same *block pattern* of symmetry.

The blocks in (3.2) are formed by making a partition $(1^b, n_1, n_2, \cdots, n_h)$ of the t rows and t columns of $\|A^{ij}\|$. A matrix having the block pattern of symmetry of (3.2) will be called *block symmetric of type I*. Clearly a block symmetric matrix of type I is invariant over all permutations of its rows and columns within the subsets determined by $(1^b, n_1, \cdots, n_h)$, if the row interchanges and column interchanges are the same. Also, a t-component vector will be called *block symmetric* if the order of values of the components is invariant over all permutations of the components within groups determined by $(1^b, n_1, \cdots, n_h)$.

The determinant of the block symmetric matrix $\|A_{ij}\|$ is

$$|A_{ij}| = K\prod_{1}^{h}(A_a - B_a)^{n_a - 1}, \tag{3.3}$$

where

$$K = \left|\begin{array}{cccc:cccc} & & & & C'_{11} & C'_{12} & \cdots & C'_{1h} \\ & A_{ss'} & & & \cdot & & \cdot & \cdot \\ & & & & \cdot & & & \\ & & & & C'_{b1} & C'_{b2} & \cdot & C'_{bh} \\ \hdashline C'_{11} & C'_{21} & \cdot & C'_{b1} & A'_1 & D'_{12} & \cdots & D'_{1h} \\ C'_{12} & C'_{22} & \cdots & C'_{b2} & D'_{21} & A'_2 & \cdots & D'_{2h} \\ \cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\ & \cdot & \cdot & \cdot & & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & & \cdot & & \cdot \\ C'_{1h} & C'_{2h} & \cdot & C'_{bh} & D'_{h1} & D'_{h2} & \cdot & A'_h \end{array}\right|,$$

[1] In general a chance quantity and the variable of its distribution function will be denoted by the same symbol.

$$
(3.2)\qquad \| A^{ij} \| = \left\| \begin{array}{cccc:cccc:cccc:c:cccc}
A^{11} & A^{12} & \cdots & A^{1b} & C^{11} & C^{11} & \cdots & C^{11} & C^{12} & C^{12} & \cdots & C^{12} & & C^{1h} & C^{1h} & \cdots & C^{1h} \\
A^{21} & A^{22} & \cdots & A^{2b} & C^{21} & C^{21} & \cdots & C^{21} & C^{22} & C^{22} & \cdots & C^{22} & \cdots & C^{2h} & C^{2h} & \cdots & C^{2h} \\
\vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots & & \vdots & & \ddots & \vdots \\
A^{b1} & A^{b2} & \cdots & A^{bb} & C^{b1} & C^{b1} & \cdots & C^{b1} & C^{b2} & C^{b2} & \cdots & C^{b2} & & C^{bh} & C^{bh} & \cdots & C^{bh} \\ \hdashline
C^{11} & C^{21} & \cdots & C^{b1} & A^{1} & B^{1} & \cdots & B^{1} & D^{12} & D^{12} & \cdots & D^{12} & & D^{1h} & D^{1h} & \cdots & D^{1h} \\
C^{11} & C^{21} & \cdots & C^{b1} & B^{1} & A^{1} & \cdots & B^{1} & D^{12} & D^{12} & \cdots & D^{12} & \cdots & D^{1h} & D^{1h} & \cdots & D^{1h} \\
\vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots & & \vdots & & \ddots & \vdots \\
C^{11} & C^{21} & \cdots & C^{b1} & B^{1} & B^{1} & \cdots & A^{1} & D^{12} & D^{12} & \cdots & D^{12} & & D^{1h} & D^{1h} & \cdots & D^{1h} \\ \hdashline
C^{12} & C^{22} & \cdots & C^{b2} & D^{21} & D^{21} & \cdots & D^{21} & A^{2} & B^{2} & \cdots & B^{2} & & D^{2h} & D^{2h} & \cdots & D^{2h} \\
C^{12} & C^{22} & \cdots & C^{b2} & D^{21} & D^{21} & \cdots & D^{21} & B^{2} & A^{2} & \cdots & B^{2} & \cdots & D^{2h} & D^{2h} & \cdots & D^{2h} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots & & \vdots & \vdots & \ddots & \vdots \\
C^{12} & C^{22} & \cdots & C^{b2} & D^{21} & D^{21} & \cdots & D^{21} & B^{2} & B^{2} & \cdots & A^{2} & & D^{2h} & D^{2h} & \cdots & D^{2h} \\ \hdashline
 & & \vdots & & & \vdots & & & & \cdot & & & \ddots & & & \vdots & \\ \hdashline
C^{1h} & C^{2h} & \cdots & C^{bh} & D^{h1} & D^{h1} & \cdots & D^{h1} & D^{h2} & D^{h2} & \cdots & D^{h2} & \cdots & A^{h} & B^{h} & \cdots & B^{h} \\
C^{1h} & C^{2h} & \cdots & C^{bh} & D^{h1} & D^{h1} & \cdots & D^{h1} & D^{h2} & D^{h2} & \cdots & D^{h2} & \cdot & B^{h} & A^{h} & \cdots & B^{h} \\
\vdots & \vdots & & \vdots & \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots & & \vdots & & \ddots & \vdots \\
C^{1h} & C^{2h} & \cdots & C^{bh} & D^{h1} & D^{h1} & \cdots & D^{h1} & D^{h2} & D^{h2} & \cdots & D^{h2} & \cdot & B^{h} & B^{h} & \cdots & A^{h}
\end{array} \right\|,
$$

where $C'_{sa} = C_{sa}\sqrt{n_a}$, $A'_a = A_a + (n_a - 1)B_a$, and $D'_{aa'} = D_{aa'}\sqrt{n_a n_{a'}}$ ($s = 1, \cdots, b; a, a' = 1, \cdots, h; a \neq a'$); $A_{ss'}$, C_{sa}, A_a, B_a, and $D_{aa'}$ are the cofactors of $A^{ss'}$, C^{sa}, A^a, B^a, and $D^{aa'}$, respectively in (3.2). The ellipsoid, defined by $A_{ij}(X_i - m_i)(X_j - m_j) = r_0$ (r_0 fixed and > 0), has $(n_a - 1)$ axes of equal length $(a = 1, \cdots, h)$, and each of the remaining q axes is inclined to the coordinate axes so that its direction cosines have the same block symmetry as the set of diagonal elements in (3.2).

When any of the $\bar{H}$ hypotheses is true, we represent $\| G^{ij} \|$ by $\| \bar{A}^{ij} \|$ (also $\| G_{ij} \|$ by $\| \bar{A}_{ij} \|$) which can be written as

$$
\text{(3.4)} \qquad \| \bar{A}^{ij} \| =
$$

$$
\left\|\begin{array}{cccc|cccc|c|cccc}
\bar{A}^1 & \bar{B}^1 & \cdot & \bar{B}^1 & \bar{C}^{12} & \bar{D}^{12} & \cdot & \bar{D}^{12} & & \bar{C}^{1h} & \bar{D}^{1h} & \cdot & \bar{D}^{1h} \\
\bar{B}^1 & \bar{A}^1 & \cdots & \bar{B}^1 & \bar{D}^{12} & \bar{C}^{12} & \cdots & \bar{D}^{12} & & \bar{D}^{1h} & \bar{C}^{1h} & & \bar{D}^{1h} \\
\vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots & \cdots & \vdots & & \ddots & \vdots \\
\bar{B}^1 & \bar{B}^1 & \cdots & \bar{A}^1 & \bar{D}^{12} & \bar{D}^{12} & & \bar{C}^{12} & & \bar{D}^{1h} & \bar{D}^{1h} & \cdots & \bar{C}^{1h} \\
\hline
\bar{C}^{21} & \bar{D}^{21} & \cdots & \bar{D}^{21} & \bar{A}^2 & \bar{B}^2 & \cdots & \bar{B}^2 & & \bar{C}^{2h} & \bar{D}^{2h} & \cdot & \bar{D}^{2h} \\
\bar{D}^{21} & \bar{C}^{21} & \cdots & \bar{D}^{21} & \bar{B}^2 & \bar{A}^2 & \cdots & \bar{B}^2 & & \bar{D}^{2h} & \bar{C}^{2h} & \cdots & \bar{D}^{2h} \\
\vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots & \cdots & \vdots & & \ddots & \vdots \\
\bar{D}^{21} & \bar{D}^{21} & \cdots & \bar{C}^{21} & \bar{B}^2 & \bar{B}^2 & \cdots & \bar{A}^2 & & \bar{D}^{2h} & \bar{D}^{2h} & \cdots & \bar{C}^{2h} \\
\hline
 & \vdots & & & & \vdots & & & \ddots & & \vdots & & \\
\hline
\bar{C}^{h1} & \bar{D}^{h1} & \cdots & \bar{D}^{h1} & \bar{C}^{2h} & \bar{D}^{2h} & \cdots & \bar{D}^{2h} & & \bar{A}^h & \bar{B}^h & \cdots & \bar{B}^h \\
\bar{D}^{h1} & \bar{C}^{h1} & \cdots & \bar{D}^{h1} & \bar{D}^{2h} & \bar{C}^{2h} & \cdot & \bar{D}^{2h} & & \bar{B}^h & \bar{A}^h & \cdots & \bar{B}^h \\
\vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots & \cdots & \vdots & & \ddots & \vdots \\
\bar{D}^{h1} & \bar{D}^{h1} & \cdots & \bar{C}^{h1} & \bar{D}^{2h} & \bar{D}^{2h} & \cdot & \bar{C}^{2h} & & \bar{B}^h & \bar{B}^h & \cdots & \bar{A}^h
\end{array}\right\|,
$$

where the blocks are formed by a partition (n^h) of the t rows and t columns; thus each block is an $n \times n$ array $\| \bar{A}^{ij} \|$ and $\| \bar{A}_{ij} \|$ have the same block pattern of symmetry.

A matrix having the block pattern of symmetry of (3.4) will be called *block symmetric of type II*. The determinant of $\| \bar{A}_{ij} \|$ is

$$
\text{(3.5)} \qquad | \bar{A}_{ij} | = \bar{K}^{n-1}\bar{Q},
$$

where

$$\bar{K} = \begin{vmatrix} \bar{A}_1 - \bar{B}_1 & \bar{C}_{12} - \bar{D}_{12} & \cdot\cdot & \bar{C}_{1h} - \bar{D}_{1h} \\ \bar{C}_{21} - \bar{D}_{21} & \bar{A}_2 - \bar{B}_2 & \cdot\cdot & \bar{C}_{2h} - \bar{D}_{2h} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \bar{C}_{h1} - \bar{D}_{h1} & \bar{C}_{h2} - \bar{D}_{h2} & \cdot\cdot & \bar{A}_h - \bar{B}_h \end{vmatrix},$$

$$\bar{Q} = \begin{vmatrix} \bar{A}'_1 & \bar{C}'_{12} & \cdots & \bar{C}'_{1h} \\ \bar{C}'_{21} & \bar{A}'_2 & \cdots & \bar{C}'_{2h} \\ \cdot & \cdot & & \cdot \\ \cdot & & & \\ & & & \cdot\cdot \\ \bar{C}'_{h1} & \bar{C}'_{h2} & \cdot\cdot & \bar{A}'_h \end{vmatrix},$$

where $\bar{A}'_a = \bar{A}_a + (n-1)\bar{B}_a$ and $\bar{C}'_{aa'} = \bar{C}_{aa'} + (n-1)\bar{D}_{aa'}$ $(a, a' = 1, 2, \cdots, h; a \neq a')$, $\bar{A}_a$, $\bar{B}_a$, $\bar{C}_{aa'}$, and $\bar{D}_{aa'}$ are the cofactors of $\bar{A}^a$, $\bar{B}^a$, $\bar{C}^{aa'}$, $\bar{D}'^{aa}$, respectively, in (3.4).

4. Method of obtaining the sample criteria. The probability distribution, P, of a simple, random sample, say $O_N(X_{1\alpha}, X_{2\alpha}, \cdots, X_{t\alpha})(\alpha = 1, 2, \cdots, N)$, from Π is

$$(4.1) \qquad P = \pi^{-Nt/2} | G_{ij} |^{N/2} \exp\left[-\sum_{i,j,\alpha} G_{ij}(X_{i\alpha} - m_i)(X_{j\alpha} - m_j)\right].$$

For O_N fixed, P is the likelihood function of the parameters $m_1, m_2, \cdots, m_t$, and G_{ij} $(i, j = 1, 2, \cdots, t)$. To obtain sample criteria for testing the H and $\bar{H}$ hypotheses we shall employ the Neyman-Pearson likelihood-ratio method. The details of applying this method will be given for only one of the hypotheses, since the technique of application is the same for all the hypotheses under consideration.

In applying the likelihood-ratio method we maximize P under two different sets of conditions and form the ratio of the two maxima. To derive a criterion for, say, $H_1(mvc)$, we first maximize P over the set, Ω, of admissible values of the parameters in (4.1), secondly, we maximize P over the set, ω, of admissible values of the parameters in (4.1) that satisfy $H_1(mvc)$. Let P_Ω and P_ω be these maxima, respectively. The likelihood-ratio criterion for $H_1(mvc)$ is $\lambda_1(mvc) = P_\omega/P_\Omega$; thus $0 \leq \lambda_1(mvc) \leq 1$. The sample criterion, $L_1(mvc)$, for $H_1(mvc)$ will be chosen as a single-valued function of $\lambda_1(mvc)$.

4a. Derivation of the criterion $L_1(mvc)$. The parameter spaces, Ω, and, ω, can be specified as follows:

$$\Omega \begin{cases} (1) & \| G_{ij} \| \text{ positive definite,} \\ (2) & -\infty < m_i < +\infty \, (i = 1, 2, \cdots, t), \end{cases}$$

$$\omega \begin{cases} (1) & \| A_{ij} \| \text{ positive definite and block symmetric (of type I);} \\ (2) & -\infty < m_i < +\infty, \, (m_1, m_2, \cdots, m_t) \text{ block symmetric.} \end{cases}$$

The block symmetries in $\omega(1)$ and $\omega(2)$ are for the same partition $(1^b, n_1, \cdots, n_h)$ of the t variates (see sections 2 and 3).

Maximizing P is equivalent to maximizing

$$L = \ln P = -(Nt/2) \ln \pi + (N/2) \ln |G_{ij}| - \sum_{i,j,\alpha} G_{ij}(X_{i\alpha} - m_i)(X_{j\alpha} - m_j). \tag{4.2}$$

Solving the simultaneous equations $\partial L/\partial m_i = 0 (i = 1, \cdots, t)$ and $\partial L/\partial G_{ij} = 0 (i, j = 1, \cdots, t; i \leq j)$ for m_i and G^{ij}, we have

$$\hat{m}_i = (1/N) \sum_{\alpha=1}^{N} X_{i\alpha} = \bar{X}_i,$$
$$(N/2)\hat{G}^{ij} = \sum_{\alpha=1}^{N} (X_{i\alpha} - \bar{X}_i)(X_{j\alpha} - \bar{X}_j) = v_{ij}, \tag{4.3}$$

substituting these values of the parameters into (4.1) we find that

$$P_\Omega = \pi^{-Nt/2}(2/N)^{N/2} |v_{ij}|^{-N/2} \exp[-Nt/2]. \tag{4.4}$$

In (4.3) and (4.5) each expression at the extreme right is defined by the corresponding expressions at the left.

In $\omega(2)$ there are $b + h$ groups of means, the means within a group being all equal; let m'_s be the s-th mean and m'_{r_a} be the common value of the means in the $(b + a)$-th group. Solving the simultaneous equations $\partial L/\partial m'_s = 0$, $\partial L/\partial m'_{r_a} = 0$, $\partial L/\partial A_{ss'} = 0$, $\partial L/\partial C_{sa} = 0$, $\partial L/\partial A_a = 0$, $\partial L/\partial B_a = 0$, $\partial L/\partial D_{aa'} = 0 (s, s' = 1, \cdots, b, a, a' = 1, \cdots, h; a \neq a')$, we find that

$$\hat{m}'_s = \bar{X}_s,$$
$$\hat{m}'_{r_a} = (1/Nn_a) \sum_{\alpha, i_a} X_{i_a\alpha} = \bar{X}'_{r_a}, \tag{4.5}$$

$$(N/2)\hat{A}^{ss'} = \sum_{\alpha=1}^{N} (X_{s\alpha} - \bar{X}_s)(X_{s'\alpha} - \bar{X}_{s'}) = v_{ss'},$$
$$(N/2)\hat{C}^{sa} = (1/n_a) \sum_{\alpha, i_a} (X_{s\alpha} - \bar{X}_s)(X_{i_a} - \bar{X}'_{r_a}) = u'_{sa},$$
$$(N/2)\hat{A}^{a} = (1/n_a) \sum_{\alpha, i_a} (X_{i_a\alpha} - \bar{X}_{r_a})^2 = v'_a,$$
$$(N/2)\hat{B}^{a} = [1/n_a(n_a - 1)] \sum_{\alpha, i_a, j_a} (X_{i_a\alpha} - \bar{X}'_{r_a})(X_{j_a\alpha} - \bar{X}'_{r_a}) = w'_a,$$
$$(N/2)\hat{D}^{aa'} = (1/n_a n_{a'}) \sum_{\alpha, i_a, i_{a'}} (X_{i_a\alpha} - \bar{X}'_{r_a})(X_{i_{a'}\alpha} - \bar{X}'_{r_a}) = z'_{aa'},$$

where $i_a, j_a = b + \bar{n}_a + 1, \cdots, b + \bar{n}_{a+1}; i_a \neq j_a; \bar{n}_a = n_1 + \cdots + n_{a-1}$; $\bar{n}_1 = 0; a, a' = 1, \cdots, h; a \neq a'$.

When $H_1(mvc)$ is true, the maximum likelihood estimates of m_i, σ_i, and $\rho_{ij}(i, j = 1, \cdots, t)$ would be obtained by means of (4.5) and the definition of $\| A^{ij} \|$ given just after (3.1).

Substituting the expressions in (4.5) into (4.1) we find that

$$P_\omega = \pi^{-Nt/2} |v'_{ij}|^{-N/2} (2/N)^{N/2} \exp[-Nt/2], \tag{4.6}$$

where

$$
(4.7) \qquad \| v'_{ij} \| = \left\|
\begin{array}{cccc:cccc:cccc:c:cccc}
v_{11} & v_{12} & \cdots & v_{1b} & u'_{11} & u'_{11} & \cdots & u'_{11} & u'_{12} & u'_{12} & \cdots & u'_{12} & & u'_{1h} & u'_{1h} & \cdots & u'_{1h} \\
v_{21} & v_{22} & \cdots & v_{2b} & u'_{21} & u'_{21} & \cdots & u'_{21} & u'_{22} & u'_{22} & \cdots & u'_{22} & & u'_{2h} & u'_{2h} & \cdots & u'_{2h} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \cdots & \vdots & \vdots & & \vdots \\
v_{b1} & v_{b2} & \cdots & v_{bb} & u'_{b1} & u'_{b1} & \cdots & u'_{b1} & u'_{b2} & u'_{b2} & \cdots & u'_{b2} & & u'_{bh} & u'_{bh} & \cdots & u'_{bh} \\
\hdashline
u'_{11} & u'_{21} & \cdots & u'_{b1} & v'_{1} & w'_{1} & \cdots & w'_{1} & z'_{12} & z'_{12} & \cdots & z'_{12} & & z'_{1h} & z'_{1h} & \cdots & z'_{1h} \\
u'_{11} & u'_{21} & \cdots & u'_{b1} & w'_{1} & v'_{1} & \cdots & w'_{1} & z'_{12} & z'_{12} & \cdots & z'_{12} & & z'_{1h} & z'_{1h} & \cdots & z'_{1h} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \cdots & \vdots & \vdots & & \vdots \\
u'_{11} & u'_{21} & \cdots & u'_{b1} & w'_{1} & w'_{1} & \cdots & v'_{1} & z'_{12} & z'_{12} & \cdots & z'_{12} & & z'_{1h} & z'_{1h} & \cdots & z'_{1h} \\
\hdashline
u'_{12} & u'_{22} & \cdots & u'_{b2} & z'_{21} & z'_{21} & \cdots & z'_{21} & v'_{2} & w'_{2} & \cdots & w'_{2} & & z'_{2h} & z'_{2h} & \cdots & z'_{2h} \\
u'_{12} & u'_{22} & \cdots & u'_{b2} & z'_{21} & z'_{21} & \cdots & z'_{21} & w'_{2} & v'_{2} & \cdots & w'_{2} & & z'_{2h} & z'_{2h} & \cdots & z'_{2h} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \cdots & \vdots & \vdots & & \vdots \\
u'_{12} & u'_{22} & \cdots & u'_{b2} & z'_{21} & z'_{21} & \cdots & z'_{21} & w'_{2} & w'_{2} & \cdots & v'_{2} & & z'_{2h} & z'_{2h} & \cdots & z'_{2h} \\
\hdashline
 & & & & & & \vdots & & & & \vdots & & \ddots & & & \vdots & \\
\hdashline
u'_{1h} & u'_{2h} & \cdots & u'_{bh} & z'_{h1} & z'_{h1} & \cdots & z'_{h1} & z'_{h2} & z'_{h2} & \cdots & z'_{h2} & & v'_{h} & w'_{h} & \cdots & w'_{h} \\
u'_{1h} & u'_{2h} & \cdots & u'_{bh} & z'_{h1} & z'_{h1} & \cdots & z'_{h1} & z'_{h2} & z'_{h2} & \cdots & z'_{h2} & & w'_{h} & v'_{h} & \cdots & w'_{h} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \cdots & \vdots & \vdots & & \vdots \\
u'_{1h} & u'_{2h} & \cdots & u'_{bh} & z'_{h1} & z'_{h1} & \cdots & z'_{h1} & z'_{h2} & z'_{h2} & \cdots & z'_{h2} & & w'_{h} & w'_{h} & \cdots & v'_{h}
\end{array}
\right\| .
$$

From (4.4) and (4.6) it follows that the likelihood-ratio criterion for $H_1(mvc)$ is:

$$\lambda_1(mvc) = [\,|\,v_{ij}\,|\,/\,|\,v'_{ij}\,|\,]^{(N/2)}, \qquad (i, j = 1, \cdots, t) \tag{4.8}$$

Finally, as the sample criterion for $H_1(mvc)$ we choose

$$L_1(mvc) = [\lambda_1(mvc)]^{(2/N)} = [\,|\,v_{ij}\,|\,/\,|\,v'_{ij}\,|\,] \tag{4.9}$$

4b. *Preliminary calculations for evaluation of moments of $L_1(mvc)$* The determinant $|\,v'_{ij}\,|$ in (4.9) is block symmetric From (3.2), (3.3), and (4.9) it follows that:

$$L_1(mvc) = |\,v_{ij}\,|\left[\prod_{a=1}^{h}(v'_a - w'_a)^{-(n_a-1)}\right]|\,v''_{rr'}\,|^{-1}, \tag{4.10}$$

where

$$\begin{aligned} v''_{ss'} &= v_{ss'}\,; \\ v''_{sr_a} &= u'_{sa}\sqrt{n_a}\,, \\ v''_{r_ar_a} &= v'_a + (n_a - 1)w'_a\,; \\ v''_{r_ar_{a'}} &= \sqrt{n_an_{a'}}\,z'_{aa'}\,, \end{aligned}$$

$(s, s' = 1, \cdots, b; r, r' = 1, \cdots, b + h; r_a = b + a; a = 1, \cdots, h)$

Let $Y_{i\alpha} = X_{i\alpha} - m_i$ and $\bar{Y}_i = (1/N)\sum_{\alpha=1}^{N} Y_{i\alpha}$, $(i = 1, \cdots, t)$ Clearly $v_{ij} = \sum_{\alpha=1}^{N}(Y_{i\alpha} - \bar{Y}_i)(Y_{j\alpha} - \bar{Y}_j)$. When $H_1(mvc)$ is true, u'_{sa}, v'_a, w'_a, and $z'_{aa'}$, in $L_1(mvc)$, can be expressed exactly as they are expressed in (4.5) with Y substituted for X, and $(v'_a - w'_a)$ and $v''_{rr'}$ in (4.10) can be expressed as follows:

$$\begin{aligned} v'_a - w'_a = (1/n_a)\,\Big\{\sum_{i_a} v_{i_ai_a} - [1/(n_a - 1)]\sum_{i_a\neq j_a} v_{i_aj_a}\Big\} \\ + (N/n_a)\sum_{i_a}\bar{Y}^2_{i_a} - [N/n_a(n_a - 1)]\sum_{i_a\neq j_a}\bar{Y}_{i_a}\bar{Y}_{j_a}\,; \end{aligned}$$

$$\begin{aligned} v''_{ss'} &= v_{ss'}\,; \\ v''_{sr_a} &= (1/\sqrt{n_a})\sum_{i_a} v_{si_a}\,, \\ v''_{r_ar_a} &= (1/n_a)\sum_{i_a, j_a} v_{i_aj_a}\,; \\ v''_{r_ar_{a'}} &= (1/\sqrt{n_an_{a'}})\sum_{i_a, j_{a'}} v_{i_aj_a}\,. \end{aligned} \tag{4.11}$$

From (4.10) and (4.11) it follows that when $H_1(mvc)$ is true, each element of the determinants on which $L_1(mvc)$ depends consists of: (a) a quadratic form in $\bar{Y}_i$ and a linear function of the v_{ij}, or (b) merely a linear function of the

$$v_{ij} \qquad (i, j = 1, \cdots, t).$$

The joint probability density function of the v_{ij} and $\bar{Y}_i$ is

$$f(v_{ij})g(\bar{Y}_1, \cdots, \bar{Y}_t), \tag{4.12}$$

where

$$f(v_{ij}) = \frac{|G_{ij}|^{(N-1)/2}\,|v_{ij}|^{(N-t-2)/2}\exp[-\sum_{i,j} G_{ij}v_{ij}]}{\pi^{t(t-1)/4}\,\Gamma\left(\frac{N-1}{2}\right)\Gamma\left(\frac{N-2}{2}\right)\cdots\Gamma\left(\frac{N-t}{2}\right)},$$

($\|G_{ij}\|$ positive definite; $N > t$), which is the Wishart distribution [9, p. 120], and

$$g(\bar{Y}_1, \cdots, \bar{Y}_t) = |G_{ij}|^{1/2}N^{t/2}\pi^{-t/2}\exp[-N\sum_{i,j} G_{ij}\bar{Y}_i\bar{Y}_j] = g(\bar{Y}), \quad \text{say},$$

which is a normal t-variate distribution. The d-th moment ($d = 0, 1, \cdots$) of $L_1(mvc)$, when $H_1(mvc)$ is true, is

$$(4.13)\qquad E[L_1(mvc)]^d = \int_R f(v_{ij})g(\bar{Y})\,|v_{ij}|^d\,|v''_{rr'}|^{-d} \cdot \prod_{a=1}^{h}(v'_a - w'_a)^{-d(n_a-1)}\left(\prod_1^t d\bar{Y}_1\right)\prod_{i\geq j} dv_{ij},$$

where the domain, R, of integration is $-\infty < \bar{Y}_i < +\infty$ $\|v_{ij}\|$ positive semi-definite ($i, j = 1, \cdots, t$). The integral in (4.13) is evaluated in section 6 (by means of Wilks' moment-generating operators) for the case where $H_1(mvc)$ is true.

5. Remarks on Wilks' Moment-Generating Operators. Wilks' operators are applicable to a far wider class of problems than those treated in this paper The following discussion is confined to a special use of the operators.

From (4.12) it follows that

$$(5.1)\qquad \int_{R'} \frac{|v_{ij}|^{(N-t-2)/2}\exp[-\sum_{i,j} G_{ij}v_{ij}]\prod_{i\geq j} dv_{ij}}{\pi^{t(t-1)/4}\prod_{i=1}^{t}\Gamma[(N-i)/2]} = |G_{ij}|^{-(N-1)/2},$$

where R' is the region in the space of v_{ij} for which $\|v_{ij}\|$ is positive definite, and $\|G_{ij}\|$ is positive semi-definite (Of course, the probability that $\|v_{ij}\|$ is not positive definite is 0.) Let $G'_{ij} = G_{ij} + \beta_{ij}(i, j = 1, \cdots, t)$; if all the β_{ij} are sufficiently small, $\|G'_{ij}\|$ is positive definite, and we have

$$(5.2)\qquad |G_{ij}|^{(N-1)/2}\int_{R'} \frac{|v_{ij}|^{N-t-2/2}\exp[-\sum_{i,j} G_{ij}v_{ij}]\prod_{i\geq j} dv_{ij}}{\pi^{t(t-1)/4}\prod_{i=1}^{t}\Gamma[(N-i)/2]} = |G_{ij}|^{(N-1)/2}\,|G'_{ij}|^{-(N-1)/2},$$

which is $E(g)$, where $g = \exp[-\sum_{i,j}\beta_{ij}v_{ij}]$.

Let I^1_{ij} be an operator (whose operand is a function of all the β_{ij}) which represents the following set of operations: (a) replacement of each β_{ij} in the operand

by $B_{ij} + \xi_i\xi_j$, (b) integration (of the result of (a)) with respect to $\xi_i (i = 1, \cdots, t)$ from $-\infty$ to $+\infty$, (c) multiplication of the result of (a) and (b) by $\pi^{-t/2}$. From (3.1) it follows that

$$(5.3) \qquad I^1_{ij}(g) = I^1_{ij}\,(\exp[-\sum_{i,j} \beta_{ij} v_{ij}]) = g \mid v_{ij} \mid^{-1/2}, \qquad (\| v_{ij} \| \text{ pos. def.});$$

and if all the β_{ij} are set equal to 0 after performing the I-operations, then $g = 1$ and (5.3) yields $| v_{ij} |^{-1/2}$. Let I^λ_{ij} be λ repetitions ($\lambda = 1, 2, \cdots$) of I^1_{ij}. Clearly,

$$(5.4) \qquad E[I^\lambda_{ij}(g)] \,|_{\beta_{ij}=0} = E[g \mid v_{ij} \mid^{-\lambda/2}] \,|_{\beta_{ij}=0} = E[| v_{ij} |^{-\lambda/2}].$$

Under all conditions of their use in this paper the I operations are interchangeable with the E operation [8; p. 316]; thus,

$$E[I^\lambda_{ij} g] = I^\lambda_{ij}[E(g)]$$

From (5.2), (5.4), and [8, pp. 318–320] we have

$$(5.5) \qquad \begin{aligned} E[| v_{ij} |^{-\lambda/2}] &= | G_{ij} |^{(N-1)/2} \{ I^\lambda_{ij} \mid G'_{ij} \mid^{-(N-1)/2} \} \,|_{\beta_{ij}=0} \\ &= | G_{ij} |^{\lambda/2} \prod_{i=1}^{t} \psi[N - i, -\lambda], \end{aligned}$$

where $N \geq t + \lambda + 1$ and $\psi(R, S) = \left[\Gamma\left(\frac{R+S}{2}\right)\right] \Big/ \left[\Gamma\left(\frac{R}{2}\right)\right]$.

The operator I_{ij} may be used, as indicated above, to find negative half-integer moments of $| v_{ij} |$. To obtain positive half-integer moments of $| v_{ij} |$ we may use an inverse operator $I^{-\lambda}_{ij}$ [8, pp. 321–323] ($\lambda = 1, 2, \cdots$) which has been defined in such a way that

$$(5.6) \qquad \begin{aligned} | G_{ij} |^{(N-1)/2} \,|\, \{ I^{-\lambda}_{ij} \mid G'_{ij} \mid^{-(N-1)/2} \} \,|_{\beta_{ij}=0} &= E[| v_{ij} |^{\lambda/2}] \\ &= | G_{ij} |^{-\lambda/2} \left(\prod_{i=1}^{t} \psi[N - i, \lambda] \right). \end{aligned}$$

The equality between the second and third expressions in (5.6) can be obtained from (5.1) by replacing N by $N + \lambda$ (see [7]).

In (5.5) and (5.6) the β's are not necessary; however, in (4.13) and in similar expressions for the moments of the other criteria there are several determinants, each determinant requires a distinct I-operator, and it is of great convenience to introduce a distinct set of β's for each I operator. The β's associated with a given operator may initially appear in more than one of the determinants in the operand. The order in which several I-operators are used is illustrated in the following case for two:

$$(5.7) \qquad [I^\lambda_{ij} \mid G'_{ij} \mid^{-k'} (I^{-\rho}_{ij} \mid G''_{ij} \mid^{-k''}) \,|_{\beta''_{ij}=0}] \,|_{\beta'_{ij}=0}\,.$$

where $\lambda, \rho > 0$ and the values of k' and k'' are such that the value of the expression is well defined. The notation in (5.7) means that $I^{-\rho}_{ij}$ is applied to $| G''_{ij} |^{-k''}$, the β's associated with $I^{-\rho}_{ij}$ are set equal to zero, then I^λ_{ij} is applied to the product

of $| G'_{ij} |^{-h'}$ and the results of the previous operations, and then the β's associated with I^{λ}_{ij} are set equal to zero. The interchangeability of the order of I operations is discussed in [8, p. 324].

6. The moments and distribution of L_1 (*mvc*) when H_1 (*mvc*) is true. To evaluate (4.13) we let

$$g = \exp\left[-\sum_{i,j} \beta_{ij} v_{ij} - \sum_a \beta'_a(v'_a - w'_a) - \sum_{r,r'} \beta''_{rr'} v''_{rr'}\right]. \tag{6.1}$$

From (4.11) and (4.12) we have

$$E(g) = | A_{ij} |^{N/2} \, | A'_{ij} |^{-(N-1)/2} \, | A''_{ij} |^{-1/2}, \tag{6.2}$$

where

$$\begin{aligned}
A'_{ss'} &= A_{ss'} + \beta_{ss'} + \beta''_{ss'}, \\
A'_{si_a} &= C_{sa} + \beta_{si_a} + \beta''_{sr_a}/\sqrt{n_a}, \\
A'_{i_ai_a} &= A_a + \beta_{i_ai_a} + \beta'_a/n_a + \beta''_{r_ar_a}/n_a, \\
A'_{i_aj_a} &= B_a + \beta_{i_aj_a} - \beta'_a/(n_a - 1)n_a + \beta''_{r_ar_a}/n_a, \qquad (i_a \neq j_a), \\
A'_{i_aj_{a'}} &= D_{aa'} + \beta_{i_aj_{a'}} + \beta''_{r_ar_{a'}}/\sqrt{n_a n_{a'}}, \qquad (a \neq a'), \\
A''_{ss'} &= A_{ss'}, \\
A''_{si_a} &= C_{sa}, \\
A''_{i_ai_a} &= A_a + \beta'_a/n_a, \\
A''_{i_aj_a} &= B_a - \beta'_a/n_a(n_a - 1), \qquad (i_a \neq j_a), \\
A''_{i_aj_{a'}} &= D_{aa'}, \qquad (a \neq a').
\end{aligned}$$

When $H_1(mvc)$ is true, we have

$$\begin{aligned}
E[L_1(mvc)]^d &= | A_{ij} |^{N/2} \left\{ \prod_{a=1}^{h} I_a^{2d(n_a-1)} \, | A''_{ij} | \, [I^{2d}_{rr'} (I^{-2d}_{ij} \, | A'_{ij} |^{-(N-1/2)})_{\beta_{ij}=0}]_{\beta''_{rr'}=0} \right\}_{\beta'_a=0} \\
&= \left\{ \prod_{i=1}^{t} \psi(N - i, 2d) \right\} \left\{ \prod_{r=1}^{q} \psi(N + 2d - r, -2d) \right\} \\
&\quad \times \left\{ \prod_{a=1}^{h} \psi[(N + 2d)(n_a - 1), -2d(n_a - 1)] \right\} \left\{ \prod_{a=1}^{h} (n_a - 1)^{d(n_a-1)} \right\},
\end{aligned} \tag{6.3}$$

$$(d = 0, 1, 2, \cdots; N > t),$$

where $q = b + h$ and $\psi(R, S)$ is defined in (5.5). In (6.3) the assumption that $H_1(mvc)$ is true implies that after we apply I^{-2d}_{ij} and set the β_{ij} equal to 0 all remaining determinants are block symmetric; we may then use (3.3) before

applying $I_{\tau\tau'}^{2d}$ and $I_a^{2d(n_a-1)}$, $(a = 1, \cdots, h)$ The expression in (6 3) may be written as follows:

$$E[L_1(mvc)]^d$$

$$= \left\{\prod_{i=1+q}^{t} \frac{\Gamma\left(\frac{N-i}{2}+d\right)}{\Gamma\left(\frac{N-i}{2}\right)}\right\}\left\{\prod_{a=1}^{h} \frac{\Gamma\left(\frac{N(n_a-1)}{2}\right)}{\Gamma\left(\frac{N(n_a-1)}{2}+d(n_a-1)\right)}\right\}$$

$$(6.4) \qquad \left(\prod_{a=1}^{h}(n_a-1)^{d(n_a-1)}\right)$$

$$= \prod_{a=1}^{h}\prod_{s_a=1}^{n_a-1}\left\{\frac{\left(\frac{N-q-s_a-\bar{n}_a+a-1}{2}\right)_d}{\left(\frac{N}{2}+\frac{(s_a-1)}{(n_a-1)}\right)_d}\right\},$$

where $\bar{n}_a$ is defined in (4 5) and $(T)_d = \Gamma(T+d)/\Gamma(T)$.

We now consider the problem of identifying from (6 4) the distribution of $L_1(mvc)$ (when $H_1(mvc)$ is true). Let θ be a beta variate, i.e, a variate whose c.d f, $F(\theta)$, is

$$(6.5) \qquad F(\theta) = I_\theta(P, Q), \qquad (0 \leq \theta \leq 1; P, Q > 0),$$

which is the Incomplete Beta Function ratio. $I_\theta(P, Q)$ is tabulated in [1] and [3]. The d-th moment of θ is:

$$(6.6) \qquad E(\theta)^d = \frac{\Gamma(P+d)}{\Gamma(P)}\frac{\Gamma(P+Q)}{\Gamma(P+Q+d)} = (P)_d/(P+Q)_d,$$

$(d = 0, 1, \cdots)$. Let

$$(6.7) \qquad \tau = \prod_{j=1}^{c}\theta_j \qquad (c = 1, 2, \cdot),$$

where the $\theta_j (j = 1, \cdots, c)$ are mutually independent and each θ_j is a beta variate, having parameters p_j, q_j, say The d-th moment of τ is

$$(6.8) \qquad E(\tau)^d = \prod_{j=1}^{c}(p_j)_d/(p_j+q_j)_d, \qquad (d = 0, 1, \cdot)$$

Given a variate, say μ $(0 \leq \mu \leq 1)$, whose d-th moment $(d = 0, 1, \cdot)$ is given by (6.8) we can infer by means of the solution of the Hausdorff problem of moments that μ and τ have the same exact probability distribution function (see Corollary 1.1 [2, p. 11]). It should be noted that (6 4) can be written as

$$(6.9) \qquad E[L_1(mvc)]^d = \prod_{a=1}^{h}\prod_{s_a=1}^{r_a-1}[(p_{as_a})_d/(p_{as_a}+q_{as_a})_d],$$

where
$$p_{as_a} = [(N - q - s_a - \bar{n}_a + a - 1)/2] > 0,$$
$$q_{as_a} = \left[\frac{(s_a - 1)}{(n_a - 1)} + \frac{q + s_a + \bar{n}_a - a + 1}{2}\right] > 0;$$
thus (6.4) is a special case of (6.8).

The exact probability (density) function, say $g(\tau)$, of τ has been obtained by Wilks [7, p. 475] and is:

$$\begin{aligned}
g(\tau) = {} & K\tau^{p_c-1}(1-\tau)^{\zeta_c-\eta_c-1}\int_0^1 \cdots \int_0^1 v_1^{q_1-1}v_2^{q_2-1} \cdots v_{c-1}^{q_{c-1}-1} \\
& \times (1-v_1)^{\zeta_{c-1}-\eta_{c-1}-1}(1-v_2)^{\zeta_{c-2}-\eta_{c-2}-1} \cdots (1-v_{c-1})^{\zeta_1-\eta_1-1} \\
& \times [1 - v_1(1-\tau)]^{p_1-p_2-q_2}[1 - \{v_1 + v_2(1-v_1)\}(1-\tau)]^{p_2-p_3-q_3} \cdots \qquad (6.10)\\
& \times [1 - \{v_1 + v_2(1-v_1) + \cdots + v_{c-1}(1-v_1)(1-v_2)\cdots(1-v_{c-2})\} \\
& \qquad (1-\tau)]^{p_{c-1}-p_c-q_c} \\
& \times \prod_{j=1}^{c-1} dv_j,
\end{aligned}$$

where $K = \prod_{j=1}^{c}\left[\frac{\Gamma(p_j + q_j)}{\Gamma(p_j)\Gamma(q_j)}\right]$, $\quad \zeta_j = \sum_{j'=0}^{j-1}(p_{c-j'} + q_{c-j'})$,

$\eta_j = \sum_{j'=0}^{j-1} p_{c-j'}$. An approximation of the distribution of a product of independent beta variates by the distribution of a single beta variate is given in [4].

The results of this section may be summarized as follows: *If $H_1(mvc)$ is true, the d-th moment ($d = 0, 1, \cdots$) of the exact distribution of $L_1(mvc)$ is given by* (6.4). Also, if $H_1(mvc)$ is true, the exact distribution of $L_1(mvc)$ is given by (6.10), where the p_j, q_j, and c can be specified by means of (6.4). The cumulative distribution of $L_1(mvc)$ is given for certain special cases in section 7g.

7. Single Sample Criteria. The solutions of problems (i) and (ii) (see section 1) for $H_1(mvc)$ are contained in (4.9) and the summary at the end of section 6. In the present section solutions of problems (i) and (ii) are given for each of the remaining two H_1 hypotheses and the three $\bar{H}_1$ hypotheses (all of which are stated in section 2). For any of the hypotheses the sample criterion is chosen as a single-valued function of the likelihood-ratio criterion for the hypothesis. The methods of determining the moments and identifying the distribution of each sample criterion (when the corresponding null hypothesis is true) are entirely similar to those used in sections 4, 5, and 6 in regard to $H_1(mvc)$. Section 7g gives the exact distributions of the single-sample criteria for certain special compound symmetries.

Each criterion discussed in this section is based on a sample

$$O_N(X_{1\alpha}, X_{2\alpha}, \cdots, X_{t\alpha})(\alpha = 1, \cdots, N, N > t)$$

of size N from a normal t-variate distribution ($t = 3, 4, \cdots$). As in the case of $H_1(mvc)$, it is presupposed for testing $H_1(vc)$ or $H_1(m)$ that there is a certain partition $(1^b, n_1, n_2, \cdots, n_h)$ of the t-variates, for testing $\bar{H}_1(mvc)$, $\bar{H}_1(vc)$, or $\bar{H}_1(m)$ it is presupposed that there is a certain partition (n^h) of the t variates (see sections 2 and 3).

7a. *The test $L_1(vc)$ for the hypothesis $H_1(vc)$.* For the sample criterion for $H_1(vc)$ we choose

$$L_1(vc) = [\lambda_1(vc)]^{2/N} = |v_{ij}| / |\bar{v}_{ij}|, \qquad (i, j = 1, \cdots, t) \tag{7.1}$$

where $\lambda_1(vc)$ is the likelihood-ratio criterion for $H_1(vc)$, v_{ij} is defined in (4.3), and

$$\begin{aligned}
\bar{v}_{ss'} &= v_{ss'}, \\
\bar{v}_{sj_a} &= (1/n_a) \sum_{j_a} v_{sj_a}, \\
\bar{v}_{i_ai_a} &= (1/n_a) \sum_{j_a} v_{j_aj_a}, \\
\bar{v}_{i_aj_a} &= [1/n_a(n_a - 1)] \sum_{i_a' \neq j_a'} v_{i_a'j_a'}, \\
\bar{v}_{i_aj_{a'}} &= (1/n_a n_{a'}) \sum_{i_a', j_{a'}'} v_{i_a'j_{a'}'},
\end{aligned}$$

($s, s' = 1, \cdots, b$; $a, a' = 1, \cdots, h$; $a \neq a'$; $i_a, i_a', j_a, j_a' = b + \bar{n}_a + 1, \cdots, b + \bar{n}_{a+1}$, $\bar{n}_a = n_1 + \cdots + n_{a-1}$; $\bar{n}_1 = 0$). Since $\| \bar{v}_{ij} \|$ is a block symmetric matrix, there is an expression for $| \bar{v}_{ij} |$ that is entirely similar in form to the expression in (3.3) for $| A_{ij} |$ (see also (4.9) and (4.10)).

If $H_1(vc)$ is true,

$$\begin{aligned}
E[L_1(vc)]^d &= \left\{\prod_{i=1}^{t} \psi(N - i, 2d)\right\} \\
&\quad \left\{\prod_{a=1}^{h} \psi[(N - 1 + 2d)(n_a - 1), -2d(n_a - 1)]\right\} \\
&\quad \times \left\{\prod_{r=1}^{q} \psi[N - r + 2d, -2d]\right\} \left\{\prod_{a=1}^{h} (n_a - 1)^{d(n_a-1)}\right\} \\
&= \prod_{a=1}^{h} \prod_{s_a=1}^{n_a-1} \left\{\frac{\left(\dfrac{N - q - s_a - n_a + a - 1}{2}\right)_d}{\left(\dfrac{N-1}{2} + \dfrac{(s_a - 1)}{(n_a - 1)}\right)_d}\right\}, \qquad (d = 0, 1, \cdots),
\end{aligned} \tag{7.2}$$

where $q = b + h$ and $\psi(R, S)$, $\bar{n}_a$ and $(T)_d$ are defined in (5.5), (4.5), and (6.4), respectively. From (7.2) and the argument given after (6.8) it follows that if $H_1(vc)$ is true, the exact distribution of $L_1(vc)$ is given by (6.10), where the p_j, q_j, and c can be specified by means of (7.2).

7b. *The test $L_1(m)$ for the hypothesis $H_1(m)$.* For the sample criterion for $H_1(m)$ we choose

$$L_1(m) = [\lambda_1(m)]^{2/N} = \frac{|\bar{v}_{ij}|}{|v'_{ij}|}, \qquad (i, j = 1, \cdots, t), \tag{7.3}$$

where $\lambda_1(m)$ is the likelihood-ratio criterion for $H_1(m)$ and v'_{ij} and $\bar{v}_{ij}$ are defined in (4.7) and (7.1), respectively. In passing we note that

$$[L_1(m)][L_1(vc)] = L_1(mvc). \tag{7.4}$$

If $H_1(m)$ is true,

$$\begin{aligned} E[L_1(m)]^d &= \prod_{a=1}^{h} \{\psi[(N-1)(n_a-1), 2a(n_a-1)] \\ &\quad \times \psi[(n_a-1)(N+2a), -2a(n_a-1)]\} \\ &= \prod_{a=1}^{h} \prod_{s_a=1}^{r_a-1} \left\{ \frac{\left(\frac{N-1}{2} + \frac{s_a-1}{n_a-1}\right)_d}{\left(\frac{N}{2} + \frac{s_a-1}{n_a-1}\right)_d} \right\}, \qquad (d = 0, 1, \cdots). \end{aligned} \tag{7.5}$$

If $H_1(m)$ is true, the exact distribution of $L_1(m)$ is given by (6.10), where the p_j, q_j and c can be specified by means of (7.5). It follows from (7.5) that the exact distribution of $L_1(m)$, when $H_1(m)$ is true, does not depend on b

7c *The test $\bar{L}_1(mvc)$ for the hypothesis $\bar{H}_1(mvc)$.* The sample criterion, $\bar{L}_1(mvc)$, for $\bar{H}_1(mvc)$ (see section 2) is

$$\bar{L}_1(mvc) = [\bar{\lambda}_1(mvc)]^{2/N} = |v_{ij}| / |\bar{v}'_{ij}|, \qquad (i, j = 1, \cdots, t) \tag{7.6}$$

where $\bar{\lambda}_1(mvc)$ is the likelihood-ratio criterion for $\bar{H}_1(mvc)$, v_{ij} is defined in (4.3), and

$$\bar{v}'_{i_a i_a} = (1/n) \sum_{\alpha, j_a} (X_{j_a\alpha} - \bar{X}'_a)^2,$$

$$\bar{v}'_{i_a j_a} = [1/n(n-1)] \sum_{\substack{\alpha \\ i'_a \neq j'_a}} (X_{i'_a\alpha} - \bar{X}'_a)(X_{j'_a\alpha} - \bar{X}'_a), \qquad (i_a \neq j_a),$$

$$\bar{v}'_{i_a k_{a'}} = (1/n) \sum_{\alpha, j_a, k'_{a'}} (X_{j_a\alpha} - \bar{X}'_a)(X_{k'_{a'}\alpha} - \bar{X}'_{a'}),$$

$$(k'_{a'} = j_a + n(a' - a);\ a \neq a'),$$

$$\bar{v}'_{i_a h_{a'}} = [1/n(n-1)] \sum_{\alpha, j_a, h'_{a'}} (X_{j_a\alpha} - \bar{X}'_a)(X_{h'_{a'}\alpha} - \bar{X}'_{a'}),$$

$$(h'_{a'} \neq j_a + n(a - a'),\ a \neq a'),$$

$(a = 1, \cdots, h, i_a, j_a, h_a, k_a = (a-1)n + 1, \cdots, an;\ k_{a'} = i_a + n(a' - a)$, $h_{a'} \neq i_a + n(a' - a);\ \alpha = 1, \cdots, N)$. $\|\bar{v}'_{ij}\|$ is a block symmetric matrix, of type II (see (3.4)), in which the blocks are formed by a partition (n^h) $(t = nh)$ of the rows and columns; there is an expression for $|\bar{v}'_{ij}|$ that is entirely similar in form to the expression in (3.5) for $|\bar{A}_{ij}|$

If $\bar{H}_1(mvc)$ is true,

$$
(7.7)\quad
\begin{aligned}
E[\bar{L}_1(mvc)]^d &= (n-1)^{hd(n-1)}\left\{\prod_{i=h+1}^{t}\psi(N-i,2d)\right\} \\
&\quad\times\left\{\prod_{a=1}^{h}\psi[(N+2d)(n-1)+1-a,\,-2d(n-1)]\right\} \\
&= \prod_{a=1}^{h}\prod_{s=1}^{n-1}\left\{\frac{\left(\dfrac{N-h-s-(n-1)(a-1)}{2}\right)_d}{\left(\dfrac{N}{2}+\dfrac{1-a}{2(n-1)}+\dfrac{s-1}{n-1}\right)_d}\right\},
\end{aligned}
\qquad (d=0,1,\cdots).
$$

If $\bar{H}_1(mvc)$ is true, the exact distribution of $\bar{L}_1(mvc)$ is given by (6.10), where the p_j, q_j and c can be specified by means of (7 7)

7d. The test $\bar{L}_1(vc)$ for the hypothesis $\bar{H}_1(vc)$. The sample criterion, $\bar{L}_1(vc)$ for $\bar{H}_1(vc)$ (see section 2) is

$$
(7.8)\qquad \bar{L}_1(vc) = [\bar{\lambda}_1(vc)]^{2/N} = |v_{ij}| / |\bar{v}_{ij}| \qquad (i, j = 1, \cdots, t),
$$

where $\bar{\lambda}_1(vc)$ is the likelihood-ratio criterion for $\bar{H}_1(vc)$, v_{ij} is defined in (4.3), and

$$
\begin{aligned}
\bar{v}_{i_a i_a} &= (1/n)\sum_{j_a} v_{j_a j_a}, && \\
\bar{v}_{i_a j_a} &= [1/n(n-1)]\sum_{i'_a \neq j'_a} v_{i'_a j'_a}, && (i_a \neq j_a), \\
\bar{v}_{i_a k_{a'}} &= (1/n)\sum_{j_a, k'_{a'}} v_{j_a k'_{a'}}, && (k'_{a'} = j_a + n(a'-a);\ a \neq a'), \\
\bar{v}_{i_a h_{a'}} &= [1/n(n-1)]\sum_{j_a, h'_{a'}} v_{j_a h'_{a'}}, && (h'_{a'} \neq j_a + n(a'-a);\ a \neq a'),
\end{aligned}
$$

where the ranges of a, i_a, j_a, h_a, k_a are given in (7.6) There is an expression for $|\bar{v}_{ij}|$ which is entirely similar in form to the expression in (3 5) for $|\bar{A}_{ij}|$

If $\bar{H}_1(vc)$ is true,

$$
(7\,9)\quad
\begin{aligned}
E[\bar{L}_1(vc)]^d &= (n-1)^{hd(n-1)}\left[\prod_{i=h+1}^{t}\psi(N-i,2d)\right] \\
&\quad\times\left\{\sum_{a=1}^{h}\psi[(N-1+2d)(n-1)+1-a,\,-2d\,(n-1)]\right\} \\
&= \prod_{a=1}^{h}\prod_{s=1}^{n-1}\left\{\frac{\left(\dfrac{N-h-s-(n-1)(a-1)}{2}\right)_d}{\left(\dfrac{N-1}{2}+\dfrac{1-a}{2(n-1)}+\dfrac{s-1}{n-1}\right)_d}\right\}, \qquad (d=0,1,\cdots).
\end{aligned}
$$

If $\bar{H}_1(vc)$ is true, the exact distribution of $\bar{L}_1(vc)$ is given by (6.10), where the p_j, q_j and c can be specified by means of (7 9).

7e. The test $\bar{L}_1(m)$ for the hypothesis $\bar{H}_1(m)$. The sample criterion $\bar{L}_1(m)$, for $\bar{H}_1(m)$ (see section 2) is

$$\bar{L}_1(m) = [\bar{\lambda}_1(m)]^{2/N} = \frac{\bar{L}_1(mvc)}{\bar{L}_1(vc)} = \frac{|\bar{v}_{ij}|}{|\bar{v}'_{ij}|}, \tag{7.10}$$

where $\bar{\lambda}_1(m)$ is the likelihood-ratio criterion for $\bar{H}_1(m)$ and $\| \bar{v}_{ij} \|$ and $\| \bar{v}'_{ij} \|$ are given in (7.8) and (7.6), respectively.

If $\bar{H}_1(m)$ is true, the d-th moment ($d = 0, 1, \cdots$) of $\bar{L}_1(m)$ is

$$E[\bar{L}_1(m)]^d = \prod_{a=1}^{h} \prod_{s=1}^{n-1} \left\{ \frac{\left(\frac{N-1}{2} + \frac{1-a}{2(n-1)} + \frac{s-1}{n-1}\right)_d}{\left(\frac{N}{2} + \frac{1-a}{2(n-1)} + \frac{s-1}{n-1}\right)_d} \right\}, \tag{7.11}$$

$$(d = 0, 1, \cdots)$$

If $\bar{H}_1(m)$ is true, the exact distribution of $\bar{L}_1(m)$ is given by (6.10) where the p_j, q_j and c can be specified by means of (7.11).

7f. Relations among $L_1(mvc)$, $L_1(vc)$, and $L_1(m)$ and among $\bar{L}_1(mvc)$, $\bar{L}_1(vc)$, and $\bar{L}_1(m)$. $L_1(mvc)$ is the product of $L_1(vc)$ and $L_1(m)$ (see (7.4)); moreover, when $H_1(mvc)$ is true, the d-th moment ($d = 0, 1, \cdots$) of $L_1(mvc)$ equals the product of the d-th moments of $L_1(vc)$ and $L_1(m)$ (see (6.4), (7.2), and (7.5)). From this result and the argument given after (6.8) it follows that when $H_1(mvc)$ is true, $L_1(mvc)$ is the product of two independent chance quantities, namely, $L_1(vc)$ and $L_1(m)$. Similarly, when $\bar{H}_1(mvc)$ is true, $\bar{L}_1(mvc)$ is the product of two independent chance quantities, namely, $\bar{L}_1(vc)$ and $\bar{L}_1(m)$.

7g. Exact distributions of single sample criteria in special cases. For a sample of size N and a partition $(1^b, n_1, \cdots, n_h)$ of the t variates of Π (see section 2) let the cumulative distribution function (c.d.f.) of $L_1(mvc)$, when $H_1(mvc)$ is true, be

$$F(u \mid 1^b, n_1, \cdots, n_h \mid N) = \text{Prob}\,\{L_1(mvc) \leq u\}; \tag{7.12}$$

also, let $F(y \mid 1^b, n_1, \cdots, n_h \mid N)$ and $F(z \mid 1^b, n_1, \cdots, n_h \mid N)$ be the c.d.f.'s of $L_1(vc)$ and $L_1(m)$ when $H_1(vc)$ and $H_1(m)$ are true, respectively. Let $F(\bar{u} \mid n^h \mid N)$, $F(\bar{y} \mid n^h \mid N)$, and $F(\bar{z} \mid n^h \mid N)$ be the c.d.f.'s of $\bar{L}_1(mvc)$, $\bar{L}_1(vc)$, and $\bar{L}_1(m)$ when $\bar{H}_1(mvc)$, $\bar{H}_1(vc)$ and $\bar{H}_1(m)$ are true, respectively.

It can be shown that

$$F(u \mid 1^b, 2 \mid N) = I_u[(N-b-2)/2, (b+2)/2],$$

$$F(u \mid 1^b, 3 \mid N) = I_{\sqrt{u}}[N-b-3, b+3],$$

$$F(y \mid 1^b, 2 \mid N) = I_y[(N-b-2)/2, (b+1)/2],$$

$$
(7.13)\quad
\begin{aligned}
F(y \mid 1^b, 3 \mid N) &= I_{\sqrt{y}}[N - b - 3, b + 2], \\
F(z \mid 1^b, n \mid N) &= I_{z'}[(N - 1)(n - 1)/2, (n - 1)/2], [z' = z^{1/(n-1)}], \\
F(\bar{u} \mid 2^2 \mid N) &= I_{\sqrt{\bar{u}}}[N - 4, 3], \\
F(\bar{y} \mid 2^2 \mid N) &= I_{\sqrt{\bar{y}}}[N - 4, 2], \\
F(\bar{z} \mid n^2 \mid N) &= I_{\bar{z}'}[(N - 1)(n - 1) - 1, n - 1], \qquad [\bar{z}' = \bar{z}^{1/2(n-1)}],
\end{aligned}
$$

where $I_x(P, Q)$ is defined in (6.5).

Distributions of the criteria in certain cases where the normal distribution is *completely symmetric* (see section 2) are given in [5].

7h *Asymptotic distributions of the single sample criteria* When the sample size, N, is large, we may use a theorem [6] (see also [9, pp. 151–2]) concerning the approximate distribution of the likelihood-ratio criterion. For large N the distributions of the quantities $-N \ln L_1(mvc)$, $-N \ln L_1(vc)$, and $-N \ln L_1(m)$ (when $H_1(mvc)$, $H_1(vc)$, and $H_1(m)$, respectively, are true) are approximately chi-square distributions with $(1/2)\ [t(t + 3) - b(b + 3) - h(h + 5)] - hb$, $(1/2)[t(t + 1) - b(b + 1) - h(h + 3)] - hb$, and $t - b - h$ degrees of freedom, respectively. Also, for large N the distributions of the quantities $-N \ln \bar{L}_1(mvc)$, $-N \ln \bar{L}_1(vc)$, and $-N \ln \bar{L}_1(m)$ (when $\bar{H}_1(mvc)$, $\bar{H}_1(vc)$, and $\bar{H}_1(m)$, respectively, are true) are approximately chi-square distributions with $[t(t + 3)/2 - h(h + 2)]$, $[t(t + 1)/2 - h(h + 1)]$, and $t - h$ degrees of freedom, respectively.

8. k-Sample Criteria. In this section solutions of problems (i) and (ii) (see section 1) are given for the three H_k and the three $\bar{H}_k$ hypotheses (all stated in section 2).

A test of any of these hypotheses is based on k simple, random samples ($k \geq 2$) from k compound-symmetric, normal t-variate distributions. The probability density function, Q, of the k samples, say, $O_{N_p}(p = 1, \cdots, k;\ N_p > b + h)$ is

$$
(8.1)\qquad Q = \pi^{-N't/2}\left[\prod_{p=1}^{k} |\, G_{ij,p} \,|^{N_p/2}\right] \times \exp\left[-\sum_{i,j,p,a_p} G_{ij,p}(X_{ia_p} - m_{i,p})(X_{ja_p} - m_{j,p})\right],
$$

$(N' = \sum_{p=1}^{k} N_p\ ;\ i, j = 1, \cdots, t)$, where X_{ia_p} is the a_p-th sample value of the i-th variate in the p-th population ($a_p = 1, \cdots, N_p$), $m_{i,p}$ is the mean (expected value) of the i-th variate in the p-th population, and $(1/2)\ \|\, G_{ij,p} \,\|^{-1}$ is the variance-covariance matrix of the variates in the p-th population (see (3.1)). For a given set of k samples Q is the likelihood function of the parameters $G_{ij,p}$ and $m_{i,p}$ ($i, j = 1, \cdots, t;\ p = 1, \cdots, k$).

The six hypotheses under consideration (see section 2) can be restated in terms of $G_{ij,p}$ and $m_{i,p}$, e.g., $H_k(MVC \mid mvc)$ asserts that $m_{i,1} = m_{i,2} = \cdots = m_{i,k}$ and $\| G_{ij,1} \| = \| G_{ij,2} \| = \cdots = \| G_{ij,k} \|$ given that for all p the vector $(m_{1,p}, \cdots, m_{t,p})$ is block symmetric and the matrix $\| G_{ij,p} \|$ is block symmetric (of type I) for a preassigned partition $(1^b, n_1, \cdots, n_h)$ of the t variates (see sections 2 and 3)

8a *Expressions for the criteria.* Let $\lambda_k(MVC \mid mvc), \cdots, \bar{\lambda}_k(M \mid mvc)$ represent the likelihood-ratio criteria for the six hypotheses $H_k(MVC \mid mvc), \cdots, \bar{H}_k(M \mid mVC)$ respectively, and let $L_k(MVC \mid mvc), \cdots, \bar{L}_k(M \mid mVC)$ be the sample criteria for the respective hypotheses. We choose the L_k as follows:

$$
\begin{aligned}
L_k(MVC \mid mvc) &= [\lambda_k(MVC \mid mvc)]^2, \\
L_k(VC \mid mvc) &= [\lambda_k(MC \mid mvc)]^2, \\
L_k(M \mid mVC) &= [\lambda_k(M \mid mVC)]^{2/N'}, \\
&= \left\{ \frac{L_k(MVC \mid mvc)}{L_k(VC \mid mvc)} \right\}^{1/N'};
\end{aligned} \tag{8.2}
$$

the expressions for $\bar{L}_k(MVC \mid mvc)$, $\bar{L}_k(VC \mid mvc)$, and $\bar{L}_k(M \mid mVC)$ are the same as those in (8.2) with λ_k replaced by $\bar{\lambda}_k$. The λ_k and $\bar{\lambda}_k$ can be obtained explicitly by straightforward application of the likelihood-ratio method (see the paragraph preceding section 4a).

8b *Moments of the k-sample criteria* The exact distribution of any of the k-sample criteria, when the corresponding null hypothesis is true, is given in (6.10), where the quantities p_i, q_j, and c can be specified by means of the moment expressions given below. The moments have been obtained by means of the operators discussed in Section 5.

For each of the following six moment expressions the null hypothesis, corresponding to the sample criterion involved, is assumed to be true:

$$
\begin{aligned}
E[L_k(MVC \mid mvc)]^d &= \left\{ \frac{\prod_{p=1}^{k} \prod_{r=1}^{q} \prod_{u_p=1}^{N_p} \left(\frac{1}{2} - \frac{r}{2N_p} + \frac{(u_p - 1)}{N_p} \right)_d}{\prod_{r=1}^{q} \prod_{u=1}^{N'} \left(\frac{1}{2} - \frac{r}{2N'} + \frac{(u - 1)}{N'} \right)_d} \right\} \\
&\times \left\{ \frac{\prod_{p=1}^{k} \prod_{a=1}^{h} \prod_{u'_p=1}^{N_p(n_a - 1)} \left(\frac{1}{2} + \frac{(u'_p - 1)}{N'(n_a - 1)} \right)_d}{\prod_{a=1}^{h} \prod_{u'=1}^{N'(n_a - 1)} \left(\frac{1}{2} + \frac{(u' - 1)}{N'(n_a - 1)} \right)_d} \right\};
\end{aligned} \tag{8.3}
$$

$$
E[L_k(VC \mid mvc)]^d = \left\{ \frac{\prod_{p=1}^{k} \prod_{r=1}^{q} \prod_{u_p=1}^{N_p} \left(\frac{1}{2} - \frac{r}{2N_p} + \frac{(u_p - 1)}{N_p} \right)_d}{\prod_{r=1}^{q} \prod_{u=1}^{N'} \left(\frac{1}{2} - \frac{(k + r - 1)}{2N'} + \frac{(u - 1)}{N'} \right)_d} \right\}
$$

$$\times \left\{ \frac{\prod_{p=1}^{k}\prod_{a=1}^{h}\prod_{u'_p=1}^{N_p(n_a-1)}\left(\frac{1}{2}+\frac{(u'_p-1)}{N_p(n_a-1)}\right)_d}{\prod_{a=1}^{h}\prod_{u'=1}^{N'(n_a-1)}\left(\frac{1}{2}+\frac{(u'-1)}{N'(n_a-1)}\right)_d} \right\};$$

$$E[L_k(M \mid mVC)]^d = \prod_{r=1}^{q} \left\{ \frac{\left(\frac{N'-k+1-r}{2}\right)_d}{\left(\frac{N'-r}{2}\right)_d} \right\};$$

$$E[\bar{L}_k(MVC \mid mvc)]^d = \left\{ \frac{\prod_{p=1}^{k}\prod_{a=1}^{h}\prod_{u_p=1}^{N_p}\left(\frac{1}{2}-\frac{a}{2N_p}+\frac{(u_p-1)}{N_p}\right)_d}{\prod_{a=1}^{h}\prod_{u=1}^{N'}\left(\frac{1}{2}-\frac{a}{2N'}+\frac{u-1}{N'}\right)_d} \right\}$$

$$\times \left\{ \frac{\prod_{p=1}^{k}\prod_{a=1}^{h}\prod_{u'_p=1}^{N_p(n-1)}\left(\frac{1}{2}+\frac{1-a}{2N_p(n-1)}+\frac{(u'_p-1)}{N_p(n-1)}\right)_d}{\prod_{a=1}^{h}\prod_{u'=1}^{N'(n-1)}\left(\frac{1}{2}+\frac{1-a}{2N'(n-1)}+\frac{(u'-1)}{N'(n-1)}\right)_d} \right\};$$

$$E[\bar{L}_k(VC \mid mvc)]^d = \left\{ \frac{\prod_{p=1}^{k}\prod_{a=1}^{h}\prod_{u_p=1}^{N_p}\left(\frac{1}{2}-\frac{a}{2N_p}+\frac{(u_p-1)}{N_p}\right)_d}{\prod_{a=1}^{h}\prod_{u=1}^{N'}\left(\frac{1}{2}-\frac{(k+a-1)}{2N'}+\frac{(u-1)}{N'}\right)_d} \right\}$$

$$\times \left\{ \frac{\prod_{p=1}^{k}\prod_{a=1}^{h}\prod_{u'_p=1}^{N_p(n-1)}\left(\frac{1}{2}+\frac{1-a}{2N_p(n-1)}+\frac{(u'_p-1)}{N_p(n-1)}\right)_d}{\prod_{a=1}^{h}\prod_{u'=1}^{N'(n-1)}\left(\frac{1}{2}+\frac{1-a}{2N'(n-1)}+\frac{(u'-1)}{N'(n-1)}\right)_d} \right\};$$

$$E[\bar{L}_k(M \mid mVC)]^d = \prod_{a=1}^{h} \left\{ \frac{\left(\frac{N'-k+1-a}{2}\right)_d}{\left(\frac{N'-a}{2}\right)_d} \right\},$$

where $d = 0, 1, \cdots$ and $(T)_d$ is defined in (6.4).

8c. *Comments on the criteria.* By an argument similar to that used in section 7f it follows from (8 3) that when $H_k(MVC \mid mvc)$ is true $L_k(MVC \mid mvc)$ is the product of two independently distributed chance quantities, namely, $L_k(VC \mid mvc)$ and $[L_k(M \mid mVC)]^{N'}$. The same assertion holds true if we replace each L by $\bar{L}$ and H by $\bar{H}$.

Exact distributions of the k-sample criteria, when the corresponding null hypotheses are true, can be obtained explicitly for special values of k and special compound symmetries, but owing to lack of space we shall not consider them in this paper.

When the sample size N' is large, the exact distributions of

$$-\ln L_k(MVC \mid mvc), \quad -\ln L_k(VC \mid mvc), \quad -N' \ln L_k(M \mid mVC),$$
$$-\ln \bar{L}_k(MVC \mid mvc), \; -\ln \bar{L}_k(VC \mid mvc),$$

and $-N' \ln \bar{L}_k(M \mid mVC)$ (if the corresponding null hypotheses, respectively, are true) are approximately chi-square distributions with

$$(k - 1)\left[\frac{b(b + 3)}{2} + hb + \frac{h(h + 5)}{2}\right],$$
$$(k - 1)[b(b + 1)/2 + hb + h(h + 3)/2],$$

$q(k - 1)$, $h(h + 2)(k - 1)$, $h(h + 1)(k - 1)$, and $h(k - 1)$ degrees of freedom, respectively.

9. Illustrative examples. The first of the following two examples[2] illustrates the use of $L_1(mvc)$, $L_1(vc)$, and $L_1(m)$ in a psychometrics experiment, the second example illustrates the use of $\bar{L}_1(mvc)$, $\bar{L}_1(vc)$, and $\bar{L}_1(m)$ in a medical-research experiment (see section 1).

EXAMPLE 1. In an experiment to establish methods of obtaining reader reliability in regard to essay scoring, 126 examinees were given a three-part English Composition examination. Each part required that the examinee write an essay, and for each examinee four scores were obtained on the following four things, respectively: (1) the part-2 and part-3 essays together, (2) the original part-1 essay, (3) a long-hand copy of the part-1 essay, (4) a carbon copy of the long-hand copy in (3). Scores were assigned by a group of "English Readers" using procedures designed to counterbalance certain experimental conditions The score on (1) serves as a criterion. The experimenter asks whether on the basis of the sample (of size 126) the quantities associated with (2), (3), and (4) can be considered as interchangeable among themselves and interchangeable with respect to their relation to the criterion (1).

Let X_1, X_2, X_3, and X_4 be the scores on (1), (2), (3), and (4), respectively It is assumed that (X_1, X_2, X_3, X_4) has a normal 4-variate distribution and that the set of scores $(X_{1\alpha}, X_{2\alpha}, X_{3\alpha}, X_{4\alpha})$ $(\alpha = 1, \cdots, 126)$ obtained from the essays is a random sample of values of (X_1, X_2, X_3, X_4). The following three questions will be considered (see section 2), where the grouping of the four variates is (1, 3): (a) Is the sample consistent with the hypothesis $H_1(mvc)$? (b) Is the sample consistent with the hypothesis $H_1(vc)$? (c) Is the sample consistent with the hypothesis $H_1(m)$? In the particular experiment under discussion (a) is the experimenter's question.

---

[2] Mr L R. Tucker (Educational Testing Service, Princeton, New Jersey) and Captain J. Allan Rafferty, M.D. (Air University School of Aviation Medicine, Randolph Field, Texas) kindly gave the author the data for Examples 1 and 2, respectively.

The sample means and variance-covariance matrix are as follows:

	X_1	X_2	X_3	X_4
	77.8976	20.9425	23.4544	18.0384
	20.9425	25.0704	12 4363	11.7257
	23.4544	12.4363	28.2021	9 2281
	18 0384	11.7257	9.2281	22.7390
Means	28 0556	14 9048	15.4841	14 4444

This matrix is $(1/126)\,\|v_{ij}\|$ $(i, j = 1, \cdots, 4)$ (see (4.3)). The sample criteria $L_1(mvc)$, $L_1(vc)$, and $L_1(m)$ will be used to answer questions (a), (b), and (c), respectively. The values of the criteria can be computed from the values of $|v_{ij}|$, $|v'_{ij}|$, and $|\bar{v}_{ij}|$ (see (4.9), (7.1), (7 3)), where v'_{ij} is given in (4.7) and $\bar{v}_{ij}$ is given below (7.1). The $\bar{v}_{ij}(i \neq 1 \neq j)$ are evaluated by simple averaging of certain elements in $\|v_{ij}\|$. Both $|v'_{ij}|$ and $|\bar{v}_{ij}|$ have the block pattern of (3 2) and can be expressed in the simplified form of (3.3), where $h = 1$ and $n_1 = 3$; the simplified form of $|v'_{ij}|$ can also be obtained from (4.10) and (4.11) From the data above it is found that

$$L_1(mvc) = |v_{ij}| / |v'_{ij}| = .9214,$$

$$L_1(vc) = |v_{ij}| / |\bar{v}_{ij}| = .9568,$$

$$L_1(m) = |\bar{v}_{ij}| / |v'_{ij}| = .9630.$$

The second, fourth, and fifth formulas in (7.13) (for $N = 126$, $b = 1$, $n = 3$) give the distributions of $L_1(mvc)$, $L_1(vc)$, and $L_1(m)$, respectively (when the hypothesis with which the criterion is associated is true). By direct computation with expressions for the Incomplete Beta Function ratios the per cent points corresponding to the observed values of $L_1(mvc)$, $L_1(vc)$, and $L_1(m)$ are found to be .26, .49, and 09, respectively. Thus at the 5% significance level the answer to any given one of the three questions (a), (b), (c) is yes. Critical values of $L_1(mvc)$, $L_1(vc)$, and $L_1(m)$ for various significance levels can be obtained from [3] by interpolation.

EXAMPLE 2. In an experiment to study certain properties of the blood of asphyxiated dogs, the $\%CO_2$ and hematocrit of 10 asphyxiated dogs were measured four minutes and seven minutes after asphyxiation Let X_1 and X_3 be $\%CO_2$ and hematocrit four minutes after asphyxiation, respectively, and X_2 and X_4 be $\%CO_2$ and hematocrit seven minutes after asphyxiation, respectively It is assumed that (X_1, X_2, X_3, X_4) has a normal 4-variate distribution and that the set of measurements $(X_{1\alpha}, X_{2\alpha}, X_{3\alpha}, X_{4\alpha})$ $(\alpha = 1, \cdots, 10)$ obtained from the 10 dogs is a random sample of values of (X_1, X_2, X_3, X_4). The following questions will be considered, where the grouping is (2^2): (a) Is the sample consistent with the hypothesis $\bar{H}_1(mvc)$? (b) Is the sample consistent with the hypothesis $\bar{H}_1(vc)$? (c) Is the sample consistent with the hypothesis $\bar{H}_1(m)$? In the particular experiment under discussion (a) is the experimenter's question.

The sample means and sums of squares and cross-products are as follows:

	X_1	X_2	X_3	X_4
	294.916	313.908	−89 364	−69 282
	313.908	363 689	−130.422	−69 261
	−89.364	−130.422	210.356	241.688
	−69.282	−69.261	241.688	515.789
Means	50.780	53.590	41.180	43.890.

This matrix is $\| v_{ij} \|$ $(i, j = 1, \cdots, 4)$ (see (4.3)). The sample criteria $\bar{L}_1(mvc)$, $\bar{L}_1(vc)$, and $\bar{L}_1(m)$ will be used to answer questions (a), (b), and (c), respectively. The values of these criteria can be computed from the data above (see (7 6), (7 8), and (7 10)) and are found to be:

$$\bar{L}_1(mvc) = | v_{ij} | / | \bar{v}'_{ij} | = 09107,$$

$$\bar{L}_1(vc) = | v_{ij} | / | \bar{v}_{ij} | = .3259,$$

$$\bar{L}_1(m) = | \bar{v}_{ij} | / | \bar{v}'_{ij} | = .2794.$$

The sixth, seventh, and eighth formulas in (7.13) (for $N = 10$, $n = 2$) give the distributions of $\bar{L}_1(mvc)$, $\bar{L}_1(vc)$, and $\bar{L}_1(m)$, respectively (when the hypothesis with which the criterion is associated is true). From [1] it is found that the observed values of $\bar{L}_1(mvc)$, $\bar{L}_1(vc)$, and $\bar{L}_1(m)$ correspond to the 1.2, 12.4, and 6 per cent points, respectively, of the distributions referred to above. Thus at the 5% significance level the answer to questions (a) and (c) is no and to (b) is yes. The critical values of $\bar{L}_1(mvc)$, $\bar{L}_1(vc)$, and $\bar{L}_1(m)$ for various significance levels can be found from [3].

More than one of the sample criteria may be of interest in regard to a given sample (see [5] pp. 267–268). For example, in an experiment such as that described in Example 1 suppose the answer to question (a) is no. The experimenter might then consider question (b); if the answer is no, the inconsistency of the sample with $H_1(mvc)$ might be regarded as due to the variances or covariances If the answer to (b) is yes, the experimenter might then consider (c); if the answer here is no, the inconsistency of the sample with $H_1(mvc)$ might be regarded as due to the means. If, however, the answer here is yes, further study might be required to "explain" the inconsistency.

10. Acknowledgement. The author wishes to express his gratitude to Professor S. S Wilks under whose direction this paper was written. The author also wishes to thank Professor J. W. Tukey for many valuable conversations on certain mathematical aspects of the problems and to thank Dr. W. E. Kappauf, J Allan Rafferty, M.D., and Mr. L. R. Tucker for assistance on applications of the test criteria

REFERENCES

[1] K. PEARSON, *Tables of the Incomplete Beta Function*, Cambridge University Press, 1932.

[2] J. A. SHOHAT AND J. D. TAMARKIN, *The Problem of Moments*, American Mathematical Society, 1943, pp. 9–12.

[3] CATHERINE M. THOMPSON, "Table of percentage points of the incomplete beta function," *Biometrika*, Vol. 32, Part II (1941), pp. 151–181.

[4] JOHN W. TUKEY AND S. S. WILKS, "Approximation of the distribution of the product of beta variables by a single beta variable," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 318–324.

[5] S. S. WILKS, "Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 257–281.

[6] S. S. WILKS, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 60–62.

[7] S. S. WILKS, "Certain generalizations in the analysis of variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.

[8] S. S. WILKS, "Moment-generating operators for determinants of product moments in samples from a normal system," *Annals of Math.*, Vol. 35 (1934), pp. 312–340.

[9] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1943.

BRANCHING PROCESSES[1]

By T. E. Harris

Project RAND, Douglas Aircraft Company

1. Summary. This paper is concerned with a simple mathematical model for a branching stochastic process Using the language of family trees we may illustrate the process as follows. The probability that a man has exactly r sons is p_r, $r = 0, 1, 2, \cdots$. Each of his sons (who together make up the first generation) has the same probabilities of having a given number of sons of his own; the second generation have again the same probabilities, and so on. Let z_n be the number of individuals in the nth generation. We study the probability distribution of z_n. Some previous results are given in section 2, these include procedures for computing moments of z_n, and a criterion for when the family has probability 1 of dying out In sections 3 and 4 the ease is considered where the family has a non-zero chance of surviving indefinitely In this case the random variables z_n/Ez_n converge in probability to a random variable w with cumulative distribution $G(u)$. It is shown that $G(u)$ is absolutely continuous for $u \neq 0$. Results of a Tauberian character are given for the behavior of $G(u)$ as $u \to 0$ and $u \to \infty$. In section 5 some examples are given where $G(u)$ can be found explicitly; $G(u)$ is computed numerically for the case $p_1 = 0.4$, $p_2 = 0.6$. In section 6 families with probability 1 of extinction are considered. A method is given for obtaining in certain cases an expansion for the moment-generating function of the number of generations before extinction occurs. In section 7 maximum likelihood estimates are obtained for the p_r and for the expectation Ez_1, consistency in a certain sense is proved. In section 8 a brief discussion is given of the relation between two types of mathematical models for branching processes.

2. Introduction. By a branching stochastic process is meant a phenomenon of the following general type: each of an initial aggregate of objects can give rise to more objects of the same or different types, the objects produced can then produce more, and the system develops, subject to certain probability laws. Examples are the development of human or animal populations, propagation of genes, and nuclear chain reactions. The mathematical model dealt with in this paper may be thought of as representing the generation-by-generation growth of a family, the fundamental random variable being the number of individuals in the nth generation. Under certain conditions, however, this model may describe the size of a family at a sequence of points in time. This question will be touched on in section 8.

---

[1] Based on a doctoral dissertation presented to the Mathematics Department, Princeton University, June, 1947.

DEFINITION 2.1. The random variables z_n, $n = 0, 1, 2, \cdots$, will be said to represent a *simple discrete branching process* provided: $z_0 = 1$; $P(z_1 = r) = p_r$, $r = 0, 1, 2, \cdots$, with $\sum_{r=0}^{\infty} p_r = 1$; the conditional distribution of z_{n+1}, given $z_n = r$, is that of the sum of r independent random variables, each having the same distribution as z_1.

Assumptions. Throughout this paper we assume that $\sum_{r=0}^{\infty} r^2 p_r < \infty$, that at least two of the p_r are positive, and that $p_0 + p_1 < 1$.

DEFINITIONS 2.2. Let $x = Ez_1 = \Sigma r p_r$, $\sigma^2 = \mathrm{Var}(z_1) = \Sigma r^2 p_r - x^2$. Let $f(s) = \sum_{r=0}^{\infty} p_r s^r$ be the generating function of z_1 (s denotes a complex variable). Let $p_{nr} = P(z_n = r)$ and $f_n(s) = \sum_{r=0}^{\infty} p_{nr} s^r$; of course $p_{1r} = p_r$ and $f_0(s) = s$. The assumptions given above insure that the first and second derivatives $f'(s)$ and $f''(s)$ are continuous in the set consisting of the interior of the unit circle and the point $s = 1$; thus derivative notations such as $f''(1)$ are used even though $f(s)$ may not be analytic at $s = 1$. It will be seen shortly that a similar remark applies to the functions $f_n(s)$ and certain functions to be introduced later.

In the remainder of this section we shall summarize certain results, most of them are contained implicitly or explicitly in works by Fisher [1], Lotka [2], Steffensen [3], Ulam and Hawkins [4], Kolmogoroff [5], Kolmogoroff and Dmitriev [6], and Yaglom [7]; some of these references are not widely available.

From our definition, $P(z_{n+1} = k \mid z_n = j)$ is the coefficient of s^k in $[f(s)]^j$. Hence $p_{n+1,k}$ is the coefficient of s^k in $\sum_{j=0}^{\infty} p_{nj}[f(s)]^j$, whence

$$f_{n+1}(s) = f_n[f(s)] \tag{2.1}$$

Letting $n = 1, 2, \cdots$, successively, it follows that *the generating function of z_n is the* n*th functional iterate of $f(s)$* Hence

$$f_{n+1}(s) = f[f_n(s)] \tag{2.2}$$

We note that $f_n'(1) = Ez_n$, $f_n''(1) + f_n'(1) - [f_n'(1)]^2 = \mathrm{Var}(z_n)$ Differentiation of (2.1) at $s = 1$ gives $f_{n+1}'(1) = x^{n+1}$; another differentiation gives $f_{n+1}''(1) = f''(1)[f_n'(1)]^2 + f'(1)f_n''(1)$ while twofold differentiation of (2.2) gives $f_{n+1}''(1) = f''(1)f_n'(1) + [f'(1)]^2 f_n''(1)$; these two expressions for $f_{n+1}''(1)$ can be equated and solved for $f_n''(1)$, provided $x = f'(1) \neq 1$. Thus *the mean and variance of z_n are given by* $Ez_n = (Ez_1)^n = x^n$; $\mathrm{Var}(z_n) = \dfrac{\sigma^2 x^n(x^n - 1)}{x^2 - x}$, $x \neq 1$; $\mathrm{Var}(z_n) = n\sigma^2$, $x = 1$. Higher moments, if they exist, may be found by a similar process.

DEFINITION 2.3. Denote by a the smallest non-negative real root of the equation $t = f(t)$. We see that $x \leq 1$ implies $a = 1$ while $x > 1$ implies $0 \leq a < 1$, the equality $a = 0$ holding if and only if $p_0 = 0$. In no case can the half-open interval $0 \leq t < 1$ contain more than one root. It is readily seen that

$$\lim_{n \to \infty} p_{n0} = \lim f_n(0) = a \tag{2.3}$$

We thus have the well known result: *the number a is the probability of eventual extinction of the family.* The relation between a and x shows that *the probability of extinction is* 1 *if and only if* $x \leq 1$.

It is also clear that $0 \leq t < 1$ implies $\lim f_n(t) = a$; this, together with (2.3), shows that

$$\lim_{n \to \infty} p_{nr} = 0, r = 1, 2, \cdots . \tag{2.4}$$

Relation (2.4) means roughly that the family either dies out or gets very large In section 4 it will be shown that (2.4) holds uniformly in r.

DEFINITION 2.4. The random variables w_n are defined by $w_n = z_n/x^n$. Clearly $Ew_n = 1$ and $Ew_n^2 = 1 + \frac{\sigma^2}{x^2 - x}\left(1 - \frac{1}{x^n}\right)$ if $x \neq 1$.

Suppose $n > m$. Then $E(z_n z_m) = \sum_r p_{mr} E(rz_n \mid z_m = r) = \sum_r p_{mr} r^2 x^{n-m} = x^{n-m} Ez_m^2$. Thus $E(w_n w_m) = Ew_m^2$, whence

$$E(w_n - w_m)^2 = Ew_n^2 - Ew_m^2, \qquad n > m. \tag{2.5}$$

By virtue of (2.5) we obtain

THEOREM 2.1. *If $x > 1$, the random variables w_n converge in mean square, hence in probability, to a random variable w.*

For in this case $Ew_n^2 \to 1 + \frac{\sigma^2}{x^2 - x}$ as $n \to \infty$ and (2.5) shows that $E(w_n - w_m)^2 \to 0$ as n and $m \to \infty$. Theorem 2.1 is then a consequence of [8], p. 38, I

It is well known that convergence in mean square implies $Ew_n^2 \to Ew^2$ and $E(w_n - 1)^2 \to E(w - 1)^2$ whence $Ew_n \to Ew$.

Thus we have

$$Ew = 1, \qquad Ew^2 = 1 + \frac{\sigma^2}{x^2 - x}. \tag{2.6}$$

In order to study the behavior of z_n for large n when $x > 1$, we consider the distribution of w.

DEFINITIONS 2.5. $G_n(u) = P(w_n \leq u); \phi_n(s) = E(e^{w_n s}) = \int_{0-}^{\infty} e^{su}\, dG_n(u)$.

DEFINITIONS 2.6. (Applicable when $x > 1$.) $G(u) = P(w \leq u)$; $\phi(s) = E(e^{ws}) = \int_{0-}^{\infty} e^{su}\, dG(u)$. We shall refer to $G(u)$ as *the asymptotic distribution branching from $f(s)$.*

The moment-generating functions (m.g f.'s) $\phi_n(s)$ and $\phi(s)$ are defined at least for Re $(s) \leq 0$ Unless specifically stated otherwise we shall consider them only in that domain.

From (2.2) and the fact that $\phi_n(s) = f_n[e^{s/x^n}]$ it follows that $\phi_{n+1}(sx) = f[\phi_n(s)]$. Theorem 2.1 implies that if $x > 1$ $G_n(u) \to G(u)$ and $\phi_n(s) \to \phi(s)$ for Re $(s) \leq 0$. Thus *the m g f. $\phi(s)$ satisfies the functional equation*

$$\phi(sx) = f[\phi(s)], \qquad \text{Re } (s) \leq 0. \tag{2.7}$$

Equation (2.7), which of course is applicable only when $x > 1$, was obtained in a different form by Ulam and Hawkins It belongs to a type usually known as Koenigs' equation, after the nineteenth century mathematician who studied it in connection with functional iteration, and is related to an equation studied by Abel. We shall make some use of the work of Koenigs later. See Hadamard [9] and Koenigs [10].

We note that $Ew^k < \infty$ if and only if $Ez_1^k < \infty$. It was already pointed out that $Ew = 1$. As pointed out in [4], as many further moments of w as exist may be found by successive differentiation of (2.7) at $s = 0$.

Finally we note that $G_n(0) = p_{n0}$. Hence $\lim G_n(0) = a$ Thus $G(0) = P(w = 0) \geq a$. We show later that $G(0) = a$ Clearly $G(u) = 0$ for $u < 0$.

In sections 3 and 4 we always assume $x > 1$.

3. Asymptotic properties of the moment-generating function. We first show that (2.7) uniquely determines the distribution of w. Specifically,

THEOREM 3.1. *Let $G_1(u)$ and $G_2(u)$ be distributions with equal first moments and finite second moments whose characteristic functions $\phi_1(it)$ and $\phi_2(it)$ satisfy (t is real) $\phi_r(itx) = f[\phi_r(it)]$, $r = 1, 2$. Then $G_1(u) = G_2(u)$.*

From [13], p. 27, $\phi_1(it) - \phi_2(it) = t^2\beta(t)$, where $\beta(t)$ is bounded as $t \to 0$. From (2.7), $|\phi_1(itx) - \phi_2(itx)| = |f[\phi_1(it)] - f[\phi_2(it)]| \leq x\,|\phi_1(it) - \phi_2(it)|$, since $|f'(s)| \leq x$ when $|s| \leq 1$. Hence for $t \neq 0$, $\left|\beta\left(\frac{t}{x}\right)\right| \geq x\,|\beta(t)|$ Thus $\beta(t)$ cannot be bounded near $t = 0$ unless it is identically zero, hence

$$\phi_1(it) = \phi_2(it).$$

It is clear that the requirement that $\phi(s)$ have the form $1 + s + O(s^2)$ between two rays from the origin is sufficient for the uniqueness in that domain of solutions of (2.7). On the other hand, continuous solutions can be constructed at will if the existence of a derivative near $s = 0$ is not required.

Before proceeding further, it is convenient to define three functions $k(s)$, $\psi(s)$, and $H(u)$ which are closely related to $f(s)$, $\phi(s)$, and $G(u)$ respectively. We repeat that we are considering only the case $x > 1$. See definition 2 3 for a.

DEFINITIONS 3.1. Let $k(s) = \dfrac{f[s(1-a)+a]-a}{1-a}$. Clearly $k(s)$ is a probability generating function with $k(0) = 0$, $k'(1) = f'(1) = x$, $k''(1) < \infty$. We write $k(s) = \sum_{r=1}^{\infty} q_r s^r$. We also define the iterates $k_n(s)$ by

$$k_0(s) = s, \qquad k_{n+1}(s) = k[k_n(s)].$$

DEFINITIONS 3.2. Let $H(u)$ be the asymptotic distribution branching from $k(s)$ (See Definition 2.6.) Let $\psi(s)$ be the corresponding moment-generating function We know then that $\psi(s)$ and $k(s)$ satisfy

$$\psi(sx) = k[\psi(s)] \tag{3.1}$$

In view of the uniqueness theorem we have, by direct substitution in (3.1), that $\psi(s)$ must be given by

$$\psi(s) = \frac{\phi[(1-a)s] - a}{1-a}, \tag{3.2}$$

and that $H(u)$ must be given by

$$H(u) = \frac{G\left(\frac{u}{1-a}\right) - a}{1-a}, \quad u \geq 0; \qquad H(u) = 0, \qquad u < 0. \tag{3.3}$$

We shall see later that $H(0) = 0$; i.e., that $G(0) = a$. *Therefore $H(u)$ is the conditional distribution of $(1 - a)w$, given that $w \neq 0$* Another way of stating this is as follows:

THEOREM 3 2. *The random variable w is distributed as the product of two independent random variables $w_0 \cdot w'$, where w_0 takes the values 0 and $\frac{1}{1-a}$ with probabilities a and $1 - a$ respectively while w' has the asymptotic distribution branching from $k(s)$.*

For it is directly verifiable that $\psi(s)$ is the m.g.f of $w_0 \cdot w'$.

In theorems 3.3 and 3.4 we consider the behavior of $\psi(s)$ for large $|s|$. To make for smoother reading we defer the proofs till section 9, where somewhat more general formulations are given. In section 4 the properties of $\psi(s)$ are interpreted in terms of $G(u)$.

DEFINITION 3 3. Let $\gamma = \log_x\left(\frac{1}{q_1}\right) = \log_x\left[\frac{1}{f'(a)}\right]$. (See definitions 2 3 and 3 1.) If $q_1 = 0$ (i.e., $p_0 = p_1 = 0$) we take $\gamma = \infty$.

THEOREM 3.3. *Suppose $\gamma < \infty$. Then if* Re $(s) \leq 0$ *and $s \neq 0$,*

$$\psi(s) = \frac{M(s)}{|s|^\gamma} + M_0(s). \tag{3 4}$$

$M(s)$ is continuous for $s \neq 0$; $M(s)$ and $M_0(s)$ satisfy respectively

$$M(sx) = M(s), \qquad M_0(s) = O\left(\frac{1}{|s|^{2\gamma}}\right), \qquad |s| \to \infty \tag{3.5}$$

Remarks. (See section 9 for proof.) (a). Under the conditions of the theorem $M(s)$ is real and positive when s is real and negative. (b) If $Ez_1^r < \infty$ and the conditions of the theorem hold, the rth derivative of $\psi(s)$ satisfies

$$|\psi^{(r)}(s)| = O\left(\frac{1}{|s|^{r+\gamma}}\right), \qquad |s| \to \infty. \tag{3.6}$$

(c) If $\gamma = \infty$, $\psi(s)$ and as many derivatives as exist approach 0 exponentially as $|s| \to \infty$.

We now consider the behavior of $\psi(s)$ on the positive real axis, provided it is defined there.

LEMMA 3 1 *Let $f(s)$ be analytic in the circle $|s| < \alpha$, $\alpha > 1$ Then $\phi(s)$ and $\psi(s)$ are analytic in some neighborhood of $s = 0$.*

We use a theorem of Poincaré [11] which insures that there is exactly one function $\bar{\phi}(s)$ analytic near $s = 0$ with $\bar{\phi}(0) = \bar{\phi}'(0) = 1$ and satisfying

$$\bar{\phi}(sx) = f[\bar{\phi}(s)].$$

(Although Poincaré's proof is for the case $f(s)$ rational, it applies equally well here) The circle of convergence of the MacLaurin series for $\bar{\phi}(s)$ has radius t_α where $\bar{\phi}(t_\alpha) = \alpha$. An argument whose details are given in [12], p. 21, then shows that $\phi(s) = \bar{\phi}(s)$ for $|s| < t_\alpha$, and Lemma 3.1 follows. (The argument is necessary to rule out the possibility that the $\phi_n(s)$ converge to $\bar{\phi}(s)$ for Re $(s) \leq 0$ but to some other function for Re $(s) > 0$) Clearly $\phi(s)$ and $\psi(s)$ are entire if and only if $f(s)$ is entire.

Lemma 3.1 is useful for actual computation of $G(u)$. The (non-negative) coefficients c_r in the series $\phi(s) = 1 + s + c_2s^2 + \cdots$ can be determined by differentiating (2.7) at $s = 0$. The series can be used to compute values of the characteristic function $\phi(it)$ on some interval $t_0 \leq t \leq t_0x$, where t_0 is a small real number; the values of $\phi(it)$ for the remaining values of t are determined by (2.7). (Note that the real and imaginary parts of $\phi(it)$ are respectively even and odd) Then the usual inversion formula is used to obtain $G(u)$. A numerical example of this procedure is worked out in section 5

DEFINITION 3 4. The number ρ is defined by $\rho = \log_x d$ if $f(s)$ is a polynomial of degree d, $\rho = \infty$ otherwise.

THEOREM 3 4. *Let $f(s)$ (and hence $k(s)$) be a polynomial of degree d Then for $s > 0$*

$$\frac{\log \psi(s)}{s^\rho} = L(s) + L_0(s),$$

$L(s)$ is continuous and positive, $L(s)$ and $L_0(s)$ satisfy respectively

$$L(sx) = L(s), \qquad L_0(s) = O\left(\frac{1}{s^\rho}\right), \qquad s \to \infty.$$

The proof is in section 9 (Theorem 3.4 may be compared with a more widely applicable but less precise result due to Shah [19].)

COROLLARY. *If $f(s)$ is a polynomial of degree d, $\psi(s)$ is an entire function of order ρ and type C where $C = \text{Max}\, L(s), 1 \leq s \leq x$*

An explicit determination for C has not been found An approximate numerical determination is not difficult, the function $L(s) = \lim_{n\to\infty} \frac{\log k_n[\psi(s)]}{s^\rho d^n}$ can be determined numerically for a number of values on some convenient interval $s_0 \leq s \leq s_0x$, and the maximum value approximated The importance of C will be indicated in the conjecture following Theorem 4 3. We may also mention that the quantity [Max $L(s)$ − Min $L(s)$], $1 \leq s \leq x$, is of some interest Some numerical work indicates that in certain cases $L(s)$ is at least approximately constant

4. Some properties of $G(u)$. Since it will be convenient to work with $H(u)$ rather than $G(u)$, we state the content of Theorems 4.1, 4.2, and 4.3 in terms of $G(u)$: $G(u) = a + \int_0^u g(v)\,dv$ for $u > 0$. The density $g(u)$ is continuous for $u \neq 0$. If $Ez_1^k < \infty$ then $g^{(r)}(u)$ is continuous for $u \neq 0$ provided $r < \gamma + k - 1$ and is continuous for $u = 0$ provided $r < \gamma - 1$. Near $u = 0$, $G(u)$, provided $\gamma < \infty$, approximates, in a certain mean sense made clear by Theorem 4.2, the function $a + \frac{(1-a)^{\gamma+1}}{\Gamma(1+\gamma)} u^\gamma M[u(1-a)]$, where for convenience we have defined $M(u)$ for positive u by $M(u) = M(-u)$. It is then shown that in a certain sense $g(u)$ goes to zero faster than $\exp(-u^{Q-\epsilon})$ and slower than $\exp(-u^{Q+\epsilon})$ where ϵ is any positive number, Q being defined in Theorem 4.3. A conjecture is given of a more precise result, applicable when $f(s)$ is a polynomial: in the same sense $g(u)$ goes to zero (more, less) rapidly than ($\exp[-(A^* - \epsilon)u^Q]$, $\exp[-(A^* + \epsilon)u^Q]$), where A^* is defined in the conjecture.

DEFINITION 4.1. Let $H'(u) = h(u)$.

THEOREM 4.1. *$H(u)$ is absolutely continuous.* Theorem 3.3 shows that $H(u)$ is continuous, see [13], p. 25. This incidentally shows that $G(0) = a$. If $\gamma > \frac{1}{2}$ the absolute continuity of $H(u)$ follows from the Plancherel theorem. See any text on Fourier transforms. In any case, define the functions

$$h_m(u) = \frac{1}{2\pi} \int_{-m}^{m} e^{-itu} \psi(it)\,dt, \qquad m = 1, 2, \cdots .$$

An integration by parts[2] gives for $u \neq 0$

$$h_m(u) = \frac{-1}{2\pi i u} [\psi(im)e^{-imu} - \psi(-im)e^{imu}] + \frac{1}{2\pi i u} \int_{-m}^{m} e^{-itu} \frac{d\psi(it)}{dt}\,dt. \tag{4.1}$$

If $0 < u_1 \leq u \leq u_2$, (4.1), (3.4), and (3.6) show that the continuous functions $h_m(u)$ converge uniformly in $[u_1, u_2]$ to a continuous function $h(u)$. Moreover

$$\begin{aligned} H(u_2) - H(u_1) &= \lim_{m \to \infty} \int_{-m}^{m} \frac{(e^{-itu_2} - e^{-itu_1})}{-2\pi i t} \psi(it)\,dt \\ &= \lim_{m \to \infty} \int_{u_1}^{u_2} h_m(u)\,du = \int_{u_1}^{u_2} h(u)\,du, \end{aligned} \tag{4.2}$$

the first equality in (4.2) following from [13], p. 28 and the second from the fact that the $h_m(u)$ are uniformly bounded for $u_1 \leq u \leq u_2$. In case $Ez_1^k < \infty$ and $r < \gamma + k - 1$, repeated integration by parts of (4.1) and reference to remark (b), Theorem (3.3), shows that the first r derivatives of $h(u)$ are continuous if $u \neq 0$. The usual integral expression for $h(u)$ in terms of $\psi(it)$ shows that $\gamma > r + 1$ implies $h^{(r)}(u)$ is continuous at 0.

---

[2] I am indebted to J. W. Tukey for this suggestion, which simplifies the original proof.

COROLLARY TO THE CONTINUITY OF $H(u)$ *the numbers* $p_{nr} = P(z_n = r) \to 0$ *uniformly in* r, $r \geq 1$, *as* $n \to \infty$. We have

$$p_{nr} = \left[G_n\left(\frac{r}{x^n}\right) - G\left(\frac{r}{x^n}\right)\right] + \left[G\left(\frac{r}{x^n}\right) - G\left(\frac{r}{x^n} - \frac{1}{x^n}\right)\right] + \left[G\left(\frac{r-1}{x^n}\right) - G_n\left(\frac{r-1}{x^n}\right)\right].$$

The desired result follows because $G_n(u) \to G(u)$ uniformly for $u \geq 0$ and because $G(u)$ must be uniformly continuous for $0 \leq u < \infty$ (right-continuity at 0)

We next consider the behavior of $H(u)$ near $u = 0$, when $\gamma < \infty$. Theorem 3 3 suggests what sort of result may be expected. If the function $M(s)$ of Theorem 3.3 were a constant M it would follow from a Tauberian theorem due to Karamata (see [14], pp 189–192) that $H(u) \sim \dfrac{Mu^\gamma}{\Gamma(\gamma + 1)}$ as $u \to 0+$, or $\dfrac{H(u)}{u^{\gamma+1}} \sim \dfrac{M}{u\Gamma(\gamma + 1)}$ Integrating both sides of this relation from u to ux would give

$$\int_u^{ux} \frac{H(v)\,dv}{v^{\gamma+1}} \sim \frac{1}{\Gamma(\gamma + 1)} \int_1^x \frac{M\,dv}{v} \tag{4 3}$$

The analogue of (4.3) turns out to be true, as shown by Theorem 4.2, which shows that in a certain mean sense, $H(u)$ behaves like $\dfrac{u^\gamma M(u)}{\Gamma(\gamma + 1)}$ as $u \to 0+$ (We defined $M(u) = M(-u)$ for $u > 0$)

THEOREM 4 2

$$\lim_{u \to 0+} \int_u^{ux} \frac{H(v)\,dv}{v^{\gamma+1}} = \frac{1}{\Gamma(\gamma + 1)} \int_1^x \frac{M(v)\,dv}{v}.$$

The proof, which follows directly along the lines of the proof of Karamata's theorem, is sketched briefly in section 9, for a somewhat more general situation

A corollary of Theorem 4.2 is that if $\gamma < 1$, $h(u)$ cannot be bounded as $u \to 0+$, for $h(u) < K$ implies

$$\lim_{u \to 0+} \int_u^{ux} \frac{K \cdot v dv}{\gamma^{\gamma+1}} > \frac{1}{(\gamma + 1)} \int_1^x \frac{M(v)}{v}\,dv > 0, \quad \text{or} \quad \lim_{u \to 0+} u^{1-\gamma} > 0,$$

which implies $\gamma \geq 1$. An example to be given in section 5 shows that if $\gamma = 1$, $h(u)$ is at least in certain cases bounded but discontinuous at 0

In order to consider the behavior of $H(u)$ as $u \to \infty$ we first prove a theorem which applies to any distribution whose m g.f. is an entire function.

THEOREM 4.3 [3] *Let $F(u)$ be any c d.f whose m.g f $\xi(s)$ is entire. Let ρ be the order of $\xi(s)$ Let Q be defined by*

$$Q = \text{l.u.b. } q: \int_{-\infty}^{\infty} e^{|u|^q}\,dF(u) < \infty$$

[3] Before completing the present proof, the writer communicated this result to R P. Boas, Jr , who sent back a proof along different lines

Then $\frac{1}{\rho} + \frac{1}{Q} = 1.$

The proof is given in section 9.

Combining Theorems 3.4 and 4.3, we obtain immediately

THEOREM 4.4. *Let* $Q = \text{l.u.b. } q\colon \int_0^\infty e^{u^q} h(u)\, du < \infty$. *Then* $Q = \frac{\rho}{\rho - 1}$.

Here ρ is given by definition 3.4. If $f(s)$ is not a polynomial, whether entire or not, the proof of theorem 4.3 will show that $Q = 1$, and we interpret theorem 4.4 in that sense. The trivial case $f(s) = s^k$ is excluded, so $\rho > 1$.

CONJECTURE. *Let* $\xi(s)$ *of theorem* 4.3 *be of finite order* ρ *and of type* C, $0 < C < \infty$. *Let* $Q = \frac{\rho}{\rho - 1}$ *and let* $A = \text{l.u.b. } A'\colon \int_{-\infty}^{\infty} e^{A'|u|^Q}\, dF(u) < \infty$. *Then* $(C\rho)^Q\ (AQ)^\rho = 1$.

The proof for the case ρ rational follows the same lines as the proof of Theorem 4.3; a general proof has not been found. If the conjecture is true then having determined ρ and Q, when $k(s)$ is a polynomial, and having estimated C by the procedure indicated following the corollary to theorem 3.4, we obtain

$$A = \frac{1}{Q}\left(\frac{1}{C\rho}\right)^{1/(\rho-1)} \tag{4.4}$$

for the l.u.b of the numbers A' such that $\int_0^\infty e^{A'u^Q} h(u)\, du < \infty$ The corresponding number A^* which applies to $g(u)$ is given by

$$A^* = A(1 - a)^Q. \tag{4.5}$$

5. Some special cases. In this section we shall discuss some special cases in which the m.g.f. $\phi(s)$ and the c.d.f. $G(u)$ may be determined explicitly. For these cases and for certain others there is a close relationship between the simple discrete branching process and another type of model to be discussed in section 8. Finally a numerical computation of the distribution $G(u)$ will be given for a particular case where $f(s)$ is a second degree polynomial

Suppose $f(s)$ has the form

$$f(s) = 1 - \frac{x}{\alpha} + \frac{x}{\alpha}\left(\frac{1}{1 + \alpha - \alpha s}\right)$$

with $x > 1$, $\alpha \geq x - 1$, where $f'(1) = x$ and $f''(1) + f'(1) = Ez_1^2 = x(1 + 2\alpha)$. It is easily verified (as pointed out by Poincaré in [11]), that the solution of the equation $\phi(sx) = f[\phi(s)]$ is given by $\phi(s) = 1 + \frac{(x-1)s}{x - 1 - \alpha s}$ with $\phi(0) = \phi'(0) = 1$

The number a satisfying $a = f(a)$ is given by $a = \frac{\alpha + 1 - x}{\alpha}$. The functions $\psi(s)$ and $k(s)$ of section 4 are given by $\psi(s) = \frac{1}{1 - s}$, $k(s) = \frac{s}{x - (x - 1)s}$.

The number γ of Theorem 3 3 is 1. The density function $h(u)$ (definition 4 1) is simply e^{-u}, as seen by direct calculation. The number Q of Theorem 4 3 is 1, as it should be, since $f(s)$ is not an entire function. The c.d.f $H(u)$ is $1 - e^{-u}$, and $H(u) \sim u$ near $u = 0$, in agreement with Theorem 4.2 Various aspects of the case $f(s) = \dfrac{As + B}{Cs + D}$ have been discussed by numerous authors

Somewhat more generally, we may consider generating functions of the form

$$k(s) = s[x - (x - 1)s^m]^{-1/m}, \qquad x > 1 \tag{5 1.}$$

The function $k(s)$ is a generating function if and only if m is a non-negative integer In this case we have $\phi(s) = \psi(s) = (1 - ms)^{-1/m}$ and $g(u) = h(u) = \dfrac{1}{(m^{1/m})\Gamma\left(\dfrac{1}{m}\right)} u^{(1/m)-1}e^{-(u/m)}$. Here $\gamma = \dfrac{1}{m}$, and we note that unless $m = 1$ the density function $h(u)$ is unbounded near $u = 0$. A physical interpretation for this case will be given in section 8

As a numerical illustration we consider the case $f(s) = 0.4s + 0\,6s^2$ We have $x = Ez_1 = 1\,6$ and $\sigma^2 = E(z_1 - x)^2 = 0.24$. For the asymptotic distribution, $Ew = 1, E(w - 1)^2 = \dfrac{\sigma^2}{x^2 - x} = 0\,25$. The number $\gamma = \log_{1\,6}\left(\dfrac{1}{0.4}\right) = 1.9495$ so that $\psi(s)$ which is identical with $\phi(s)$ in this case, is $O\left(\dfrac{1}{|s|^{1\,9495}}\right)$ as $|s|$ goes to ∞ with Re $(s) \leq 0$ This implies that the c.d.f $H(u)$ and likewise $G(u)$, since the two are equal here, behaves like $[1/\Gamma(1 + \gamma)]M(u)$ times $u^{1,9495}$ near $u = 0$, where the "behavior" is in the sense of Theorem 4 2. Numerical determination of $M(u)$ would not be difficult. The number ρ of Theorem 4 4 is given by $\log_x 2 = 1\,4748$. This means that $\psi(s)$ is an entire function of order 1.4748 and hence that the density function $h(u)$ goes to zero more rapidly than $e^{-u^{Q-\epsilon}}$ and less rapidly than $e^{-u^{Q+\epsilon}}$ for any $\epsilon > 0$, where $Q = \dfrac{\rho}{\rho - 1} = 3.1061$, and "more rapidly" is used in the sense of Theorem 4.4.

The function $L(s) = \lim\limits_{n\to\infty} \dfrac{\log \psi(sx^n)}{s^\rho 2^n}$ was computed for four values of s between $s = 1$ and $s = x = 1.6$, in each case the value was 0.744625 so that it appears likely that here $L(s)$ is constant Hence $C =$ Max $L(s) = 0.744625$ and the quantity A defined by (4.4) is 0.26430. Thus the conjecture following theorem 4.4 indicates that $\int_0^\infty g(u)e^{(0\,744625\pm\epsilon)u^{3\,1061}}\,du$ is (divergent, convergent) according as the + or − sign holds.

Through the kindness of Mr. Cecil Hastings of the Douglas Aircraft Company, the c d.f. $G(u)$ was computed for this case The coefficients in the power series expansion of $\phi(s)$ were obtained from the functional equation (3 1) and $G(u)$ was then obtained by inverting $\phi(it)$. The values of $G(u)$ are given in Table I.

6. Number of generations to extinction. It was pointed out in section 2 that when $x \leq 1$ the probability is 1 that $z_n = 0$ for some integer n. *We assume through-out section 6 that $x < 1$.*

TABLE I

$G(u)$, the limiting probability that $z_n/x^n \leq u$ for the case $f(s) = 0.4s + 0.6s^2$

u	$G(u)$
0.00	.00000
0 25	.04753
0.50	.17275
0.75	.34550
1.00	.53117
1.25	.69932
1 50	83042
1.75	.91857
2.00	96781
2.50	99751
3.00	.99993

DEFINITIONS 6.1 Let the random variable N be the smallest integer n such that $z_{n+1} = 0$. Define the moment-generating function of N by

$$\theta(s) = \sum_{n=0}^{\infty} e^{ns} P(N = n)$$

Clearly $P(N = n) = p_{n+1,0} - p_{n0}$, so that $\theta(s) = \sum_{n=0}^{\infty} e^{ns}(p_{n+1,0} - p_{n0})$.

DEFINITIONS 6.2. Let $b_n = 1 - p_{n+1,0}$, with $b_0 = 1 - p_0$. The numbers b_n satisfy the recursive relation

$$b_{n+1} = 1 - f(1 - b_n). \tag{6.1}$$

Define the function $\theta_1(s)$ by

$$\theta_1(s) = \sum_{n=0}^{\infty} b_n e^{ns}.$$

We see that

$$\theta(s) = 1 + (e^s - 1)\theta_1(s), \tag{6.2}$$

so that it suffices to determine the function $\theta_1(s)$.

The function $\theta_1(s)$ belongs to a type which has been studied by Fatou [15] and Lattès [16]. If we let $e^s = z$ we see that $\theta_1(z)$ is a power series whose coefficients are successive iterates of the function $f^*(b) = 1 - f(1 - b)$; i e, $b_{n+1} = f^*(b_n) = f^*_{n+1}(b_0)$, where $f^*(0) = 0$, $f^{*\prime}(0) = x < 1$. It was shown by Fatou

that a function of this sort is meromorphic with poles at $s = -n \log x$, $n = 1, 2, \cdots$. An expansion for $\theta_1(s)$ in the form

$$\theta_1(s) = \frac{\mu_1 y_0}{1 - xe^s} + \frac{\mu_2 y_0^2}{1 - x^2e^s} + \frac{\mu_3 y_0^3}{1 - x^3e^s} + \cdots,$$

was obtained by Lattès, the expansion converging everywhere except at the poles. The quantities μ_i and y_0 are defined as follows: the function $\mu(s) = \mu_1 s + \mu_2 s^2 + \mu_3 s^3 + \cdots$ is determined by the functional equation $\mu(sx) = f^*[\mu(s)]$ with the condition $\mu'(1) = \mu_1 = 1$. The number y_0 is determined by $\mu(y_0) = b_0 = 1 - p_0$. Perhaps the easiest way to determine y_0 is to use the fact that the inverse function $\mu^{-1}(s)$ satisfies the functional equation $\mu^{-1}[f^*(s)] = x\mu^{-1}(s)$, from which we can determine the power series for $\mu^{-1}(b_0)$.

Since the use of Lattès' expansion requires finding the expansions of $\mu(s)$ and $\mu^{-1}(s)$, we now give another method, giving a different kind of expansion; this method appears particularly adapted to the case here illustrated, where $f(s)$ is of the second degree. Then (6.1) becomes

$$b_{n+1} = xb_n - p_2b_n^2, \qquad b_0 = 1 - p_0 \tag{6.3}$$

DEFINITION 6.3. The functions $\theta_k(s)$, $k = 1, 2, \cdots$, are given by

$$\theta_k(s) = \sum_{n=0}^{\infty} (b_n)^k e^{ns}. \tag{6.4}$$

If we raise both sides of (6.3) to the kth power, multiply both sides by e^{ns}, sum on n from 0 to ∞, and solve for $\theta_k(s)$, we obtain

$$\theta_k(s) = \frac{b_0^k e^{-s} + \sum_{r=1}^{k} \binom{k}{r} (-p_2)^r x^{k-r} \theta_{k+r}(s)}{e^{-s} - x^k}. \tag{6.5}$$

(Justification for the rearrangement of series will come out of the subsequent proof.) If we put $k = 1$ in (6.5) we obtain

$$\theta_1(s) = \frac{b_0 e^{-s} - p_2\theta_2(s)}{e^{-s} - x}. \tag{6.6}$$

DEFINITIONS 6.4. We define recursively sequences of functions $S_n(s)$ and $R_n(s)$, such that for each n, $\theta_1(s) = S_n(s) + R_n(s)$. Let

$$S_1(s) = \frac{b_0 e^{-s}}{e^{-s} - x}, \qquad R_1(s) = -\frac{p_2\theta_2(s)}{e^{-s} - x}.$$

Suppose now that $R_n(s)$ is of the form $A_{n1}\theta_{n+1}(s) + \cdots + A_{nn}\theta_{2n}(s)$, the A_{nj} being functions of s, p_2, and x, but not explicitly of b_0, while $S_n(s)$ is a rational function of e^{-s}, p_2, and x, and a polynomial of degree n in b_0. Now put $k = n + 1$ in (6.5) and substitute the expression obtained for $\theta_{n+1}(s)$ into $R_n(s)$. Collecting terms we now define $R_{n+1}(s)$ as the sum of terms involving $\theta_{n+2}(s), \cdots, \theta_{2n+2}(s)$: $R_{n+1}(s) = A_{n+1,1}\theta_{n+2}(s) + \cdots + A_{n+1,n+1}\theta_{2n+2}(s)$; then $S_{n+1}(s) =$

$\theta_1(s) - R_{n+1}(s)$ is a rational function of e^{-s}, p_2, and x, and a polynomial of degree $n + 1$ in b_0.

THEOREM 6.1. *Let $f(s) = p_0 + p_1 s + p_2 s^2$, with $x < 1$. Suppose that $x + p_2 b_0 < 1$. Then the functions $S_n(s)$ converge to $\theta_1(s)$ in a neighborhood of $s = 0$*

The restriction $x + p_2 b_0 < 1$ may fail to hold. However this is not a serious restriction; we pick a value of n so that $x + p_2 b_n < 1$. Then

$$\theta_1(s) = b_0 + \cdots + b_{n-1}e^{(n-1)s} + e^{ns}\theta_1^*(s),$$

where $\theta_1^*(s) = \sum_{j=n}^{\infty} b_j e^{(j-n)s}$ is the same type of function as $\theta_1(s)$; theorem 6.1 is then applicable to $\theta_1^*(s)$.

If the conditions of theorem 6.1 are satisfied, we have

$$\theta_1(s) = b_0 e^{-s}[\pi_1(s, x) - p_2 b_0 \pi_2(s, x) + 2xp_2^2 b_0^2 \pi_3(s, x) - p_2^3 b_0^3(e^{-s} + 5x^3)\pi_4(s, x) + \cdots] \tag{6.7}$$

where $\pi_k(s, x) = \prod_{r=1}^{k}\left(\frac{1}{e^{-s} - x^r}\right)$. Since $E(N) = \theta'(0) = \theta_1(0)$ and $E(N^2) = \theta''(0) = 2\theta_1'(0) + \theta_1(0)$, we have

$$E(N) = b_0[\pi_1(0, x) - p_2 b_0 \pi_2(0, x) + 2xp_2^2 b_0^2 \pi_3(0, x) - p_2^3 b_0^3 (1 + 5x^3)\pi_4(0, x) + \cdots],$$

$$E(N^2) = -E(N) + 2b_0[\pi_1'(0, x) - p_2 b_0 \pi_2'(0, x) + 2xp_2^2 b_0^2 \pi_3'(0, x) - (5x^3 + 1)p_2^3 b_0^3 \pi_4'(0, x) + p_2^3 b_0^3 \pi_4(0, x) + \cdots]$$

where $\pi_k'(0, x) = \pi_k(0, x) \sum_{r=1}^{k} \frac{1}{1 - x^r}$.

We now prove that if $x + p_2 b_0 < 1$, the expansion (6.7) is valid in some neighborhood of $s = 0$ We shall denote the particular values of x, p_2, and b_0 with which we are dealing by $\bar{x}$, $\bar{p}_2$, and $\bar{b}_0$. Now let x, p_2, and b_0 be three complex numbers, arbitrary except for the following restrictions:

$$|x| + |p_2| < 1, \qquad |b_0| < 1 \tag{6.8}$$

and define the numbers b_n in terms of b_0, x, and p_2, by means of (6.3), with $\theta_k(s)$ defined by (6.4).

We first show that (6.7) is valid if (6.8) holds, and then show that the domain of validity also includes the original numbers $\bar{x}$, $\bar{p}_2$, and $\bar{b}_0$, provided

$$\bar{x} + \bar{p}_2\bar{b}_0 < 1.$$

If (6.8) is satisfied, we have $|b_n| < A|x|^n$ where A is a positive constant. Now suppose $1 < T < \frac{1}{x}$. Then the series defining $\theta_k(s)$, $k = 1, 2, \cdots$, are

uniformly and absolutely convergent in the domain $|e^s| \leq T$. Moreover, if $|x| + |p_2| = \Delta < 1$, we have $|b_n| \leq b_0\Delta^n$ whence, if k is an integer large enough so that $T\Delta^k < \frac{1}{2}$,

$$|\theta_k(s)| \leq 2b_0^k \tag{6.9}$$

for $|e^s| \leq T$. *In what follows, we assume* $|e^s| \leq T$. Now write $\theta_1(s) = S_n(s) + \sum_{j=1}^{n} A_{nj}(p_2, x, s)\theta_{n+j}(s)$, where n is large enough so that $T\Delta^n < \frac{1}{2}$ Let $A_n(p_2, x, s) = \underset{1 \leq j \leq n}{\text{Max}} |A_{nj}(p_2, x, s)|$. Passing to the next stage we see that $A_{n+1} \leq A_n + \frac{|A_{n1}|}{e^{-s} - x^n}\Delta^{n+1} \leq A_n\left(1 + \frac{\Delta^{n+1}}{e^{-s} - x^n}\right)$ Hence the numbers A_n are bounded. This fact, together with (6 9), shows that $\lim_{n\to\infty} R_n(s) = 0$.

Now suppose that x and b_0 have their original values $\bar{x}$ and $\bar{b}_0$ while p_2 is small enough in absolute value so that $\bar{x} + |p_2| < 1$. In this case $\lim_{n\to\infty} S_n(s) = \theta_1(s)$ We observe that $S_n(s)$ is a polynomial of degree $n - 1$ in p_2 and that $S_{n+1}(s)$ is obtained from $S_n(s)$ by adding a single term of degree n in p_2. Thus $\theta_1(s)$ has been expressed as a power series in p_2. Now consider $\theta_1(s)$ as a function of p_2, with $b_0 = \bar{b}_0$, $x = \bar{x}$. If $\bar{x} + \bar{b}_0|p_2| < 1$, we have $b_n = O[(\bar{x})^n]$ Thus $\theta_1(s)$ is analytic in p_2 for $|p_2| < \frac{1 - \bar{x}}{\bar{b}_0}$ and the expansion in (6 7), being a power series in p_2, must be valid when $\bar{x} + \bar{p}_2\bar{b}_0 < 1$.

7. Estimation of parameters. Until now we have assumed that the parameters p_r are known numbers. We may wish, however, to estimate them, having observed the numbers $z_1, z_2, \cdots, z_{n+1}$. In order to get simple maximum likelihood estimates for the p_r, it appears necessary to introduce certain auxiliary random variables

DEFINITIONS 7 1. Let z_{mk} be the number of individuals in the mth generation who have exactly k descendents in the $(m + 1)$st generation. Let $Z_n = 1 + z_1 + \cdots + z_n$.

THEOREM 7.1. *Maximum likelihood estimates of p_r and x, based on observed values of z_{mk} for $m \leq n$, are respectively,*

$$\hat{p}_r = \sum_{m=0}^{n} z_{mr}/Z_n, \qquad \hat{x} = (Z_{n+1} - 1)/Z_n.$$

(Note that the estimate $\hat{x}$ involves only $z_1, \cdots, z_{n+1}$.)

If z_m is fixed the joint conditional probability function of $z_{m0}, z_{m1}, \cdots$, is $\left[(z_m)!\prod_{r=0}^{\infty} p_r^{z_{mr}}\right] \Big/ \prod_{r=0}^{\infty}(z_{mr})!$. Thus the joint probability function of the z_{mr} for $m = 0, 1, \cdots, n$, and $r = 0, 1, 2, \cdots$, is given by the product of two factors, one of which is independent of the p_r, the logarithm of the other being $\sum_{r=0}^{\infty}(\sum_{m=0}^{n} z_{mr})$

$\log p_r$. The value of this expression is clearly maximized by taking $p_r = \hat{p}_r$ as given above. Since $\sum_r z_{mr} = z_m$ and $\sum_r r z_{mr} = z_{m+1}$, the quantity $\sum r\hat{p}_r$ gives $\hat{x}$ as above.

Although the estimates $\hat{p}_r$ are the same as we would obtain if we were dealing with Z_n trials from a multinomial distribution with probabilities p_r, the joint distribution of the quantities $\sum_{m=0}^{n} z_{mr}$, $r = 0, 1, \cdots$, is not multinomial. For example, if $Z_n > 1$ the probability of the event

$$\left\{\sum_{m=0}^{n} z_{m0} = Z_n,\ \sum_{m=0}^{n} z_{mr} = 0 \text{ for } r \neq 0\right\} \text{ is } 0.$$

We shall next show that the estimate $\hat{x}$ is, in a certain sense, consistent.

THEOREM 7.2. *If $x > 1$, the random variables Z_{n+1}/Z_n converge in probability to the random variable xV^* where $V^* = \frac{1}{x}$ if $w = 0$ and $V^* = 1$ if $w \neq 0$.*

If $w \neq 0$ then for all n, $z_n \neq 0$ and $1/Z_n \to 0$ as $n \to \infty$. Hence in this case $(Z_{n+1} - 1)/Z_n$ converges to x if Z_{n+1}/Z_n does. On the other hand, $P(w = 0) = a = P(z_n = 0)$ for some n, so that if $w = 0$, $Z_{n+1}/Z_n = 1$ with probability 1 for n large enough. Thus we need only show that Z_{n+1}/Z_n converges to x if $x > 1$ and $w \neq 0$.

We need the following:

LEMMA 7.1. *If $x > 1$, the random variables Z_n/x^n converge in probability to $\frac{wx}{x-1}$.*

Since

$$\frac{wx}{x-1} - \frac{Z_n}{x^n} = \frac{w}{x^{n+1}}\left(\frac{x}{x-1}\right) + \sum_{r=0}^{n} \frac{(w - w_r)}{x^{n-r}}, \tag{7.1}$$

it will be sufficient to show that $\lim_{n\to\infty} \left(\frac{x}{x-1}\right)^2 \frac{1}{x^{2n+2}} E(w^2) = 0$ and $\lim_{n\to\infty} E\left(\sum_{r=0}^{n} \frac{(w - w_r)}{x^{n-r}}\right)^2 = 0$. The truth of the first statement is obvious, since Ew^2 is finite. It follows from (2.5) that $E(w_r w_s) = Ew_r^2$ if $s > r$, $E(ww_r) = \lim_{n\to\infty} E(w_n w_r) = Ew_r^2$, whence $E(w - w_r)^2 = \frac{\sigma^2}{(x^2 - x)x^r}$ and $E[(w - w_r)(w - w_s)] = \frac{\sigma^2}{(x^2 - x)x^s}$ if $s > r$. Then

$$E\left(\sum_{r=0}^{n} \frac{(w - w_r)}{x^{n-r}}\right)^2 = \frac{1}{x^{2n}} \frac{\sigma^2}{x^2 - x}\left[\sum_{r=0}^{n} x^r + 2\sum_{s=1}^{n}\sum_{r=0}^{s-1} x^r\right],$$

and this quantity clearly approaches 0 as $n \to \infty$, proving Lemma 7.1.

Define the random variables w^* and V_n as

$$w^* = w \quad \text{when} \quad w \neq 0$$

$$w^* = 1 \quad \text{when} \quad w = 0$$

$$V_n = \frac{Z_n}{x^n} \quad \text{when} \quad z_n \neq 0$$

$$V_n = \frac{x}{x-1} \quad \text{when} \quad z_n = 0.$$

It is clear that the V_n converge in probability to $w^* \dfrac{x}{x-1}$, and we note that the c.d.f. of w^* is continuous at $w^* = 0$. Hence,

$$\lim_{n\to\infty} P\left(\left|\frac{V_{n+1}}{V_n} - 1\right| > \epsilon > 0\right) = \lim_{n\to\infty} P(V_{n+1} - V_n \pm V_n \epsilon \lessgtr 0)$$

$$= P\left(\frac{\pm w^* x\epsilon}{x-1} \lessgtr 0\right) = 0.$$

It follows, under the conditional hypothesis $w \neq 0$, that the variates $\dfrac{Z_{n+1}}{Z_n}$ converge in probability to x, since

$$\frac{Z_{n+1}}{Z_n} = x\frac{V_{n+1}}{V_n} \quad \text{when} \quad z_{n+1} \neq 0.$$

8. Continuous models. As mentioned in section 1 there are situations where it is more important to consider the number of individuals existing at a given time than the number in a given generation. Let a set of probabilities p_r be given. The question arises whether we can interpret these as probabilities that an individual will have a given number of descendents at the end of some fixed period of time We might then suppose that each individual in existence at that time has the same probabilities of having a given number of descendents at the end of the next (equal) length of time, these probabilities being independent of the age of the individual. A model of this sort might be considered in certain fission processes, if the probability of fission is independent of age It should be noted that the "descendents" of an individual may include the individual. For example, if a bacterium splits in two we may either regard it as having produced two descendents and dying, or as having produced one descendent and itself surviving

If an interpretation of this sort is to be satisfactory, interpolation in time must be possible In other words there should exist a family of functions $f_n(s)$ defined for all positive n such that $f_{n_1}[f_{n_2}(s)] = f_{n_1+n_2}(s)$; such that for each positive n, $f_n(s)$ is a probability generating function, $f_n(s) = \sum_{r=0}^{\infty} p_r(n)s^r$; and such that for

$n = 0, 1, 2, \cdots$ the functions $f_n(s)$ coincide with the iterates s, $f(s)$, $f[f(s)]$, $\cdots$. We may then interpret $f_n(s)$ as the generating function at time n. It is readily seen that in general such a family of functions will not exist. For example, if such a family exists we must have $f(s) = n$th iterate of $f_{1/n}(s)$ for arbitrarily large integral n, so that $f(s)$ cannot be a polynomial of degree ≥ 2

The functional equation $\phi(sx) = f[\phi(s)]$ shows that $f(s) = \phi[x\phi^{-1}(s)]$, whence $f_n(s) = \phi[x^n\phi^{-1}(s)]$ for integral n. The expression $\phi[x^n\phi^{-1}(s)]$ then might be taken as the definition of $f_n(s)$ for all positive n. See Hadamard, [9]. The problem of determining whether the functions so defined are a family of generating functions will be discussed in a subsequent paper. We remark, however, that if $f(s)$ has the form $\dfrac{s}{x - (x-1)s}$ considered in section 5 then the iterates $f_n(s)$ have the form $\dfrac{s}{x^n - (x^n - 1)s}$; they are clearly generating functions for all positive n, satisfying the required relation $f_{n_1}(f_{n_2}) = f_{n_1+n_2}$. Now suppose $g(s)$ is some function such that the function $f(s) = g^{-1}\left[\dfrac{g(s)}{x - (x-1)g(s)}\right]$ is a generating function for all $x > 1$, with $g(1) = 1$. As pointed out by Ulam and Hawkins, the iterates of functions $f(s)$ of this form are convenient to work with, the nth iterate being simply $g^{-1}\left[\dfrac{g(s)}{x^n - (x^n - 1)g(s)}\right]$. In addition, the requirement that $f(s)$ be a generating function for *all* $x > 1$ shows that the functions $f_n(s)$ are generating functions for all $n > 0$. The simplest function $g(s)$ which satisfies our requirements is $g(s) = s^m$, where m is any positive integer. In this case $f(s)$ has the form considered in (5.1) and $f_n(s) = s[x^n - (x^n - 1)s^m]^{-1/m}$. As $n \to 0$ we have $f_n(s) = \left(1 - \dfrac{n}{m}\log x\right)s + \dfrac{n \log x}{m} s^{m+1} + O(n^2)$. We may interpret this as follows. A particle in existence at a given time may, in a short time interval Δt, either split into $m + 1$ particles, with probability $\dfrac{\Delta t \log x}{m}$; or it may remain unaltered, with probability $1 - \dfrac{\Delta t \log x}{m}$. If it splits, each particle produced has the same chances for splitting as its parent, etc. Thus, from the results of section 5, it follows that if we begin with a single particle at time $t = 0$, the asymptotic probability density function for z_t/x^t, where z_t is the number of particles at time t, is given by $(m^{-1/m}u^{1/m-1}e^{-u/m})/\Gamma\left(\dfrac{1}{m}\right)$.

It is, of course, customary to begin with the elementary probabilities for a certain number of births in a short time Δt and determine the functions $f_n(s)$ from these by means of differential equations. See, for example, Arley, [17]. The results of the present paper can be applied in some cases to the continuous problem even when an explicit determination of the $f_n(s)$ is difficult. A discussion will be given in a later paper.

9. Some proofs. We give in this section proofs for (A) theorem 3.3, (B) theorem 3.4, (C) theorem 4.2, and (D) theorem 4.3; in certain cases we shall indicate slightly more general results.

(A) We make use of a result of Koenigs, in the form applicable here

KOENIGS' THEOREM: *If* $|s| \leq \lambda < 1$ *and* $q_1 \neq 0$, *then* $k_n(s) = q_1^n B(s) [1 + O(q_1^n)]$ *where* $B(s)$ *is analytic for* $|s| \leq \lambda$ *and satisfies the functional equation* $B[k(s)] = q_1 B(s)$.

Here, $O(q_1^n)$ means bounded by Aq_1^n, where A is independent of s. We remark that $B(s) \not\equiv 0$. The proof of Koenigs' theorem follows readily if we write $k_n(s) = q_1^{n-1} k(s) \prod_{j=1}^{n-1} \left\{1 + \frac{\xi[k_j(s)]}{q_1}\right\}$, where $\xi(s) = \frac{k(s)}{s} - q_1$.

Now let t_1 be a positive number such that $|\psi(s)| < 1$ when $0 < |s| \leq t_1$ and $\mathrm{Re}(s) \leq 0$. (For the rest of this proof we assume $\mathrm{Re}(s) \leq 0$.) Such a number exists; on the imaginary axis we have $\psi(it) = 1 + it - \frac{1}{2}E[(w')^2]t^2 + o(t^2)$ where $E[(w')^2] > 1$, w' having the distribution branching from $k(s)$, showing that $|\psi(it)| < 1$ if $t \neq 0$ and sufficiently small, while if $\mathrm{Re}(s) < 0$ we refer to the expression $\psi(s) = \int_0^\infty e^{su} dH(u)$. Let $\lambda = \mathrm{Max}\, |\psi(s)|$ for $t_1/x \leq |s| \leq t_1$. If $|s| > t_1$ let $N(s)$ be the smallest integer such that $|s|/x^{N(s)} \leq t_1$. Then $\psi(s) = k_{N(s)}[\psi(s/x^{N(s)})] = q_1^{N(s)} B[\psi(s/x^{N(s)})][1 + O(q_1^{N(s)})] = B[\psi(s)][1 + O(q_1^{N(s)})]$. Now $B(\psi(sx)) = q_1 B[\psi(s)]$. Let $M(s) = |s|^\gamma B[\psi(s)]$. Then $M(sx) = M(s)$. Also $\log_x |s/t_1| \leq N(s) < 1 + \log_x |s/t_1|$, and theorem 3.3 follows. Clearly $M(s)/|s|^\gamma$ is continuous for $t_1 x \leq |s| \leq t_1$, and hence, by functional continuation, wherever $\mathrm{Re}(s) \leq 0$, $s \neq 0$.

Concerning the remarks following Theorem 3.3 we have the following:

(a) If $Ez_1^r < \infty$, r-fold differentiation of $\psi(sx^n) = k_n[\psi(s)]$ gives, for $|s| \geq t_1 > 0$,

$$\psi^{(r)}(s) = \frac{1}{x^{nr}} \sum_{j=1}^{r} Q_{rj} k_n^{(j)} \left[\psi\left(\frac{s}{x^n}\right)\right], \tag{9.1}$$

where Q_{rj} is a polynomial in $\psi^{(1)}\left(\frac{s}{x^n}\right), \cdots, \psi^{(r)}\left(\frac{s}{x^n}\right)$. Now $|k_n(s)| = O(q_1^n)$ when $|s| \leq \lambda$; because of analyticity, the same must be true of $|k_n^{(j)}(s)|$. Put $n = N(s)$ in (9.1), $N(s)$ being the integer defined above. Since $k_N^{(j)}[\psi(s/x^N)] = O(q_1^N) = O((1/|s|^\gamma))$, remark (a) follows.

(b) $B(s)$ is clearly ≥ 0 when $s \geq 0$, hence $M(s) \geq 0$ when $s < 0$. Since $B(0) = 0$, $B(s) \neq 0$ for sufficiently small $s \neq 0$; since $\psi(s) \to 0$ as $|s| \to \infty$, $M(s) \neq 0$ for $|s|$ sufficiently large, since $M(sx) = M(s)$, remark (b) follows.

(c) If $\gamma = \infty$, i.e., $q_1 = 0$, then $k_n(s)$ goes to zero with great rapidity as $n \to \infty$, if $|s| < 1$. The general line of argument is clear.

(B) *Let $k(s)$ be a polynomial of degree $d > 1$ with real coefficients, $k(s) = q_0 + \cdots + q_d s^d$, with a non-negative double point, $k(\alpha) = a \geq 0$, and such that $k(s) > s$ when $s > \alpha$. Let $\psi(s)$ be any solution of the functional equation $\psi(ms) =$*

$k[\psi(s)]$ which is continuous for $s > 0$ and satisfies $\psi(s) > \alpha$ for $s > 0$; here m is any number > 1. Then theorem 3.4 holds, with x replaced by m.

It is not difficult to show that if $\alpha < s_1 \leq s \leq s_2$, $\lim_{j\to\infty} k_j(s) = \infty$ uniformly in s. Hence $\psi(s) \to \infty$ as $s \to \infty$. Write $R(s) = \log\left(1 + \frac{1}{q_d}\sum_{j=1}^{d} q_{d-j}s^{-j}\right)$. Then $d^{-n}\cdot \log \psi(sm^n) = d^{-n} \log k_n[\psi(s)] = (1 - d^{-n}) \log q_d/(d - 1) + \log \psi(s) + \sum_{j=1}^{n} d^{-j}R(k_{j-1}[\psi(s)])$, s being taken large enough so that $R(k_{j-1}[\psi(s)])$ is continuous. Thus, since the functions $R(k_{j-1}[\psi(s)])$ are bounded, the functions $d^{-n}\log \psi(sm^n)$ converge uniformly, for s sufficiently large, to a continuous function $L^*(s)$ satisfying $L^*(ms) = dL^*(s)$. Let $L(s) = t^{-\rho}L^*(s)$, where $\rho = \log_m d$. Theorem 3.4 now follows by an argument similar to that used to conclude theorem 3.3. (Note that $\sum_{n+1}^{\infty} d^{-j}R(k_{j-1}[\psi(s)]) = O(d^{-n})$)

(C) In order to avoid negative signs we work with the Laplace transform instead of the m.g.f.

Let $H(u)$ be nondecreasing on $(0, \infty)$ with $H(0) = 0$; let $\Psi(s) = \int_0^\infty e^{-su}\,dH(u)$ be finite for $s > 0$. Suppose $\Psi(s) = \frac{M(s)}{s^\gamma} + o\left(\frac{1}{s^\gamma}\right)$ as $s \to \infty$, where $0 < \gamma < \infty$, $M(s)$ is continuous and satisfies $M(sx) = M(s)$ for $s > 0$, x being some number > 1. Then $\lim_{u\to 0+} \int_u^{ux} \frac{H(v)}{v^\gamma}\,dv = \frac{1}{\Gamma(\gamma + 1)} \int_1^x \frac{M(v)}{v}\,dv$.

Following the lines of the proof of Karamata's theorem, we see that for any $y > 0$, $\int_y^{xy} s^{\gamma-1}\Psi(s)\,ds = D + o(1)$ as $s \to \infty$ where $D = \int_1^x \frac{M(s)}{s}\,ds$; i.e., $\int_y^{xy} s^{\gamma-1}\cdot ds\int_0^\infty e^{-su}\,dH(u) = D + o(1)$, or replacing s by $(n + 1)s$, $\int_y^{xy} s^{\gamma-1}\,ds\int_0^\infty e^{-su}e^{-nsu}\cdot dH(u) = D/(n + 1)^\gamma + o(1) = \frac{D}{\Gamma(\gamma)}\int_0^\infty e^{-s}e^{-ns}s^{\gamma-1}\,ds + o(1)$. It follows as in [14], pp. 189–192, that if $F(u)$ is any function of bounded variation in $(0, 1)$ we have

$$\lim_{y\to\infty} \int_y^{xy} s^{\gamma-1}\,ds \int_0^\infty e^{-su}\,F(e^{-su})\,dH(u) = \frac{D}{\Gamma(\gamma)} \int_0^\infty e^{-s}\,F(s^{-s})s^{\gamma-1}\,ds. \tag{9.2}$$

Let $F(e^{-s}) = e^s$ if $0 \leq s \leq 1$ and 0 otherwise. Then the theorem follows from (9.2).

(D) Theorem 4.3 is true if $F(u)$ is any bounded monotone increasing function. For simplicity we assume that $F(1) = 0$; it is readily seen that this causes no loss in generality. The proof is given for the case $1 < \rho < \infty$; it will be clear that $\rho = 1$ implies $Q = \infty$, while if $\rho = \infty$ (or if $\xi(s)$ is not entire) $Q = 1$.

Suppose m and n are positive integers such that $m/n < \rho/(\rho - 1)$. Then

$$\int_1^\infty \exp\,(u^{m/n})\,dF(u) = \sum_{r=0}^{\infty} \frac{1}{r!}\int_1^\infty u^{(mr/n)}\,dF(u) \leq n\sum_{r=0}^{\infty}\frac{[(r + 1)m]!}{(rn)!}\,c_{(r+1)m} \tag{9.3}$$

where $c_k = \frac{\xi^{(k)}(0)}{k!}$, interchange of integration and summation are justified by the positiveness of all terms involved. Suppose $0 < \epsilon < \frac{n}{m} - \left(1 - \frac{1}{\rho}\right)$; for k sufficiently large the inequality $c_k < k^{-k((1/\rho)-\epsilon)}$ is satisfied; see [18], p 253. Hence using Stirling's formula, we see that the last series in (9.3) is dominated by a series whose rth term, for r sufficiently large, is controlled by the factor $r^{rm(1-(1/\rho)+\epsilon-(n/m))}$. Since $1 - \frac{1}{\rho} + \epsilon - \frac{n}{m}$ is negative, the series, and hence the integral, converges. We have thus proved $\frac{1}{Q} + \frac{1}{\rho} \leq 1$.

Conversely, suppose $\frac{m}{n} > \frac{\rho}{\rho - 1}$. Let $\xi(s) = \sum_{k=0}^{m-1} \xi_k(s)$, where $\xi_k(s) = \sum_{r=0}^{\infty} c_{k+rm} \cdot s^{k+rn}$, $k = 0, 1, \cdots, m-1$. At least one of the functions $\xi_k(s)$ must be of order ρ We suppose that $\xi_0(s)$ is, if not the argument would need only slight modifications We have

$$\int_1^{\infty} \exp\,(u^{m/n})\, dF(u) \geq n \sum_{r=0}^{\infty} \frac{(rm)!\, c_{rm}}{[(r+1)n]!}. \tag{9.4}$$

Suppose $0 < \epsilon < 1 - \frac{1}{\rho} - \frac{n}{m}$. From [18], p 253, the inequality $c_{rm} > (rm)^{-rm(1/\rho+\epsilon)}$ must hold for infinitely many values of r. As in the first half of the proof this shows that the series and the integral in (9.4) diverge. Thus $\frac{1}{\rho} + \frac{1}{Q} \geq 1$ and the proof is complete

If ρ is rational, the conjecture following theorem 4 3 can be proved in a similar manner making use of a relation between the class of an entire function and the coefficients of its series expansion; see [14], p. 95

REFERENCES

[1] R A Fisher, *The Genetical Theory of Natural Selection*, Oxford, the Clarendon Press, 1930

[2] Alfred J Lotka, *Theorie Analytique des Associations Biologiques*, 2, Paris, Herman, 1930

[3] J. F. Steffensen, "Deux problemes du calcul des probabilites," *Annales de l'Inst. Henri Poincaré*, Vol 3 (1933), p. 331.

[4] D Hawkins and S Ulam, *Theory of Multiplicative Processes*, I, Los Alamos Declassified Document 265, 1944

[5] A Kolmogoroff, "Zur losung einer biologischen aufgabe", *Communications for Math and Mechanics at Tchebycheff University*, Tomsk, Vol 2, part 1 (1938), p 1, see *Fortschritte der Math*, Vol 64 (1938), part 2, p. 1223

[6] A Kolmogoroff and N. A Dmitriev, "Branching stochastic processes," *C R (Doklady) Acad Sci URSS (N S)*, Vol 56 (1947), pp 5–8 See *Math Rev* Vol. 9, No 1 (1948), p 46

[7] A M Yaglom, "Certain limit theorems of the theory of branching random processes," *Doklady Akad Nauk SSSR (N S)*, Vol 56 (1947), pp. 795–798 See *Math Rev* Vol. 9, No 3 (1948), p 149

[8] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Chelsea, 1946.

[9] J. HADAMARD, "Two works on iteration and related questions," *Am. Math. Soc. Bull*, Vol. 50 (1944), p. 67

[10] G. KOENIGS, "Nouvelles recherches sur les integrales de certaines equations fonctionnelles," *Annales Sci de l'Ecole Normale Sup de Paris*, Vol 1, series 3, suppl. (1884), p 3

[11] H. POINCARÉ, "Sur une classe nouvelle de transcendantes uniformes," *Journal de Math Pures et Appliquees*, Vol 6, 4th series (1890), p. 313

[12] T. E HARRIS, *Some Theorems on the Bernoullian Multiplicative Process*, Dissertation, Princeton, 1947.

[13] H CRAMÉR, *Random Variables and Probability Distributions*, Cambridge Tract 36, 1937

[14] D. V. WIDDER, *The Laplace Transform*, Princeton University Press, 1941.

[15] P. FATOU, "Sur une classe remarkable de series de Taylor," *Annales Sci de l'Ecole Normale Sup. de Paris*, Vol. 27, series 3 (1910), pp 43–53.

[16] S. LATTÈS, "Sur les suites recurrentes non lineaires et sur les fonctions generatrices de ces suites," *Annales de la Fac. des Sciences de l'Univ de Toulouse*, Vol. 3, series 3 (1911), pp. 96–105

[17] NIELS ARLEY, *On the Theory of Stochastic Processes and Their Applications to the Theory of Cosmic Radiation*, G E.C Gads Forlag, Copenhagen, 1943.

[18] E. C. TITCHMARSH, *The Theory of Functions*, 2d Ed., Oxford, 1939.

[19] S. M. SHAH, "On real continuous solutions of algebraic difference equations," *Am Math. Soc. Bull.*, Vol. 53 (1947), pp. 548–558.

MOST POWERFUL TESTS OF COMPOSITE HYPOTHESES. I. NORMAL DISTRIBUTIONS

By E. L Lehmann and C. Stein

University of California, Berkeley

Summary. For testing a composite hypothesis, critical regions are determined which are most powerful against a particular alternative at a given level of significance. Here a region is said to have level of significance ϵ if the probability of the region under the hypothesis tested is bounded above by ϵ. These problems have been considered by Neyman, Pearson and others, subject to the condition that the critical region be similar. In testing the hypothesis specifying the value of the variance of a normal distribution with unknown mean against an alternative with larger variance, and in some other problems, the best similar region is also most powerful in the sense of this paper. However, in the analogous problem when the variance under the alternative hypothesis is less than that under the hypothesis tested, in the case of Student's hypothesis when the level of significance is less than $\frac{1}{2}$, and in some other cases, the best similar region is not most powerful in the sense of this paper. There exist most powerful tests which are quite good against certain alternatives in some cases where no proper similar region exists. These results indicate that in some practical cases the standard test is not best if the class of alternatives is sufficiently restricted

1. Introduction. The problem to be discussed in this paper is that of testing a composite hypothesis against a simple alternative. More specifically let $\mathcal{F} = \{f\}$ be a family of probability density functions defined over a Euclidean space R_n and let g be a probability density function not in $\mathcal{F}$. We wish to test the hypothesis H_0 that the random variable $X = (X_1, \cdots, X_n)$ is distributed according to a density f of $\mathcal{F}$ against the alternative H_1 that X is distributed according to g By a test we mean a region of rejection, w in R_n.

Neyman and Pearson, in the fundamental paper [1] which laid the groundwork of the theory of optimum tests, restricted their considerations to similar regions. They considered a region (set) w to be optimum for the given level of significance ϵ if it maximizes the power

$$\int_w g(x)\, dx \tag{1}$$

subject to the restriction

$$\int_w f(x)\, dx = \epsilon \quad \text{for all} \quad f \text{ in } \mathcal{F}. \tag{2}$$

As Neyman, Wald and others have pointed out, it is more natural to replace the condition of similarity (2) by the weaker restriction

$$\int_w f(x)\, dx \leq \epsilon \quad \text{for all} \quad f \text{ in } \mathcal{F}. \tag{3}$$

A region w maximizing (1) subject to (3) is called most powerful against the alternative g at the level of significance Here and throughout the paper, all functions and sets are assumed to be Borel measurable.

In the present paper we shall consider certain composite hypotheses, and derive tests for them which are most powerful against a simple alternative. For the cases in which these tests coincide with the standard similar regions it will thus be established that no further increase in power is possible with tests of fixed sample sizes In the more usual situation where the most powerful test depends strongly on the specific alternative chosen, no such absolute justification of the standard test is possible. In these cases, any justification must take account of the fact that it is desired to obtain good power against a large class of alternatives. This can be done, for instance, by using Wald's definition of a most stringent test [2] or his concept of minimizing the maximum risk.[1] If, on the other hand, the class of alternatives is sufficiently restricted, the results of the present paper indicate that for small samples there may exist a test which is appreciably better than the standard test.

Frequently the probability of an error of the first kind is an analytic function of a nuisance parameter for every choice of critical region. Hence, if it is known that some nuisance parameter θ lies, say in a certain finite interval I, then any test which is similar for θ in I will be similar for all θ Consequently, the knowledge concerning θ cannot be used to find a more powerful test. On the other hand, as is indicated at the end of section 5, restrictions of the nuisance parameters may, for small samples, lead to considerably more powerful tests if the condition of similarity is replaced by the weaker condition (3).

There is one class of problems to which it may be desirable to apply the method of the present paper regardless of sample size; namely, if no similar region exists. Suppose, for instance, that $X_1, \cdots, X_n$ are known to be normally and independently distributed, X_i having unknown mean and variance ξ_i and σ_i^2 for $i = 1, \cdots, n$. For testing the hypothesis

$$H_0 : \sigma_i = 1, \qquad (i = 1, \cdots, n)$$

no similar region exists, while it is easy to see that against any simple alternative

$$H_1 \quad \sigma_i = \sigma_{i1} < 1, \qquad \xi_i = \xi_{i1},$$

there exists a test which satisfies condition (3) and which has good power against H_1 provided the σ_{i1} are sufficiently small.

The present first part of this paper is restricted to hypotheses concerning normal distributions. It is intended to extend the considerations to exponential

[1] In an unpublished paper, it is shown by G. Hunt and C. Stein that the traditional test is most stringent in several cases, including the (univariate) linear hypothesis and the hypothesis specifying the ratio of the variances of two normal distributions These results can be extended to analogous problems for distributions other than the normal, and similar results can be proved regarding minimization of the maximum risk if the weight function has a certain type of symmetry

and rectangular distributions, to consider non-parametric problems and possibly also more complicated problems connected with normal distributions, in later parts of the paper.

2. Sufficient conditions for a most powerful test. The method which will be used in this paper to obtain most powerful tests is an adaptation of the fundamental lemma of Neyman and Pearson [1]. At the same time it is essentially a special case of much more general results of Wald [3, 4], although the exact conditions of Wald's investigation are not satisfied in most of our problems.

Let h and g be two functions defined over R_n, let k be a constant and let w be a region in R_n such that

$$
\begin{aligned}
&g(x) \geq k\ h(x) \text{ in } w; \\
&g(x) \leq k\ h(x) \text{ in } R_n - w.
\end{aligned} \tag{4}
$$

Then if w' is such that

$$\int_{w'} h(x)\, dx \leq \int_{w} h(x)\, dx, \tag{5}$$

it follows as in the fundamental lemma where in (5) equality is assumed instead of inequality, that

$$\int_{w'} g(x)\, dx \leq \int_{w} g(x)\, dx. \tag{6}$$

Throughout the present paper we shall be concerned with the special case in which $\mathcal{F}$ is an s-parameter family. We may denote the members of $\mathcal{F}$ by f_θ and we shall obtain all members of $\mathcal{F}$ as θ ranges over a set ω in an s-dimensional Euclidean space. In the theorem which we shall now state, we shall be concerned with point functions λ defined over ω. We shall assume that $\lambda = c\mu$ where c is a positive constant and μ a cumulative distribution function.[2] Also we suppose that $f_\theta(x)$ is a measurable function of x and θ jointly. However, the theorem is also valid if ω is an abstract space and λ a (finite) non-negative additive set function (measure) over ω. Such more general interpretation may be required when applying the theory to non-parametric problems.

THEOREM 1. *Let H_0 be the hypothesis that the random variable X is distributed according to a density function f_θ with θ in ω, and let H_1 denote the alternative that X is distributed according to a density g. Let λ be a function defined over ω and such that*

$$\lambda = c\mu, \tag{7}$$

[2] The introduction of the distribution μ is simply a mathematical device and does not imply that θ is a random variable (see Wald [16] p 282)

where c is a positive constant and μ a cumulative distribution function. Let k be a constant and let w be a region in R_n such that

$$(8)\qquad \begin{aligned} g(x) &\geq k \int_w f_\theta(x)\, d\lambda(\theta) \quad \text{in} \quad w, \\ g(x) &\leq k \int_w f_\theta(x)\, d\lambda(\theta) \quad \text{in} \quad R_n - w. \end{aligned}$$

Suppose that w is of level of significance ϵ for testing H_0 against H_1, that is that

$$(9)\qquad \int_w f_\theta(x)\, dx \leq \epsilon \quad \text{for all} \quad \theta \text{ in } \omega,$$

and suppose that the subset of ω for which

$$(10)\qquad \int_w f_\theta(x)\, dx < \epsilon$$

has λ-measure zero. Then w is most powerful for testing H_0 against H_1 at level of significance ϵ.

Proof. Without loss of generality we shall assume $c = 1$. Let w' be any test of level of significance ϵ Then

$$(11)\qquad \int_{w'} f_\theta(x)\, dx \leq \epsilon \quad \text{for all} \quad \theta \text{ in } \omega,$$

and because of (7)

$$(12)\qquad \int_\omega \left\{\int_{w'} f_\theta(x)\, dx\right\} d\lambda(\theta) \leq \epsilon \int_\omega d\lambda(\theta) = \epsilon.$$

Since λ is of bounded variation we may interchange the order of integration in (12) and obtain

$$(13)\qquad \int_{w'} h(x)\, dx \leq \epsilon,$$

where

$$(14)\qquad h(x) = \int_\omega f_\theta(x)\, d\lambda(\theta).$$

From (9) and the condition surrounding (10) it follows that

$$(15)\qquad \int_\omega \left\{\int_w f_\theta(x)\, dx\right\} d\lambda(\theta) = \epsilon,$$

and therefore that

$$(16)\qquad \int_w h(x)\, dx = \epsilon.$$

Thus w and w' satisfy conditions (4) and (5), and hence also (6) which completes the proof.

It is useful to notice that, the assumptions of theorem 1 will be satisfied provided

$$\int_w f_\theta(x)\,dx$$

attains its maximum ϵ at all points of increase of λ, and therefore in particular whenever w is a similar region of size ϵ.

We shall in many problems exhibit a function λ which satisfies the conditions of theorem 1 without giving the reasons which led us to this function However the following comments concerning the tentative process that we used, may be helpful. One may first examine the known most powerful similar region. If there exists a cumulative distribution function λ such that (8) is the most powerful similar region, the problem is solved. If the most powerful similar region cannot even be approximated by (8) with a sequence of λ's, it is reasonable to conclude that the most powerful test is not similar Because the probability (under the null hypothesis) of any test is in all the problems considered here an analytic function of the parameter, this implies that the probability (under the null hypothesis) of the most powerful test attains its maximum at an at most denumerable (in some cases finite) set of points In all the cases of this kind which we considered in the present part I, it was then possible to prove the existence of a function λ with a single point of increase, which satisfied the conditions of theorem 1

A theorem analogous to theorem 1 holds for most powerful similar regions. Let H_0 and H_1 be as before and let λ be a function of bounded variation not necessarily non-decreasing. Let w be a region in R_n such that

$$(17)\qquad \begin{aligned} g(x) &\geq k \int_\omega f_\theta(x)\,d\lambda(\theta) \quad \text{in} \quad w; \\ g(x) &\leq k \int_\omega f_\theta(x)\,d\lambda(\theta) \quad \text{in} \quad R_n - w. \end{aligned}$$

Let w be a similar region of level of significance ϵ for testing H_0 against H_1, that is, let

$$(18)\qquad \int_w f_\theta(x)\,dx = \epsilon \quad \text{for all} \quad \theta \text{ in } \omega,$$

then w is a most powerful similar region for testing H_0 against H_1

For all the problems considered in this paper we shall prove the existence of functions λ satisfying the conditions of theorem 1, but we have not investigated the corresponding existence problem in general. On the other hand one verifies easily that for many of the cases treated here in which the most powerful test is not similar, the method for obtaining most powerful similar regions does not apply. However, for all the problems considered in the present paper the most powerful similar tests can be obtained easily by other methods [1, 5, 6, 7, 8]. For most of the problems the corresponding derivations have been carried out in the literature.

Although we restrict ourselves in the present paper to the problem of maximizing the power at a single alternative, theorem 1 clearly also applies to the more general problem of maximizing the average power over surfaces in a space of alternatives. Such problems have been considered from the point of view of similar regions by Wald, Hsu and others [9, 10, 11]

3. Testing the values of one or several variances. Let $X_1, \cdots, X_n$ be a sample from a normal population with mean ξ and variance σ^2, both unknown. We want to test the hypothesis H_0 that $\sigma = \sigma_0$ against the simple alternative that $\sigma = \sigma_1$, $\xi = \xi_1$. We shall show that the most powerful test for H_0 against H_1 is

$$\Sigma(x_i - \xi_1)^2 \leq k \quad \text{when} \quad \sigma_1 < \sigma_0, \tag{19}$$

$$\Sigma(x_i - \bar{x})^2 \geq c \quad \text{when} \quad \sigma_1 > \sigma_0, \tag{20}$$

where k and c are determined by the level of significance. Thus the best similar region is most powerful if the variance under the alternative is greater than that under the null hypothesis, while the most powerful tests against the other alternatives are not similar. That the region $\Sigma(x_i - \bar{x})^2 \geq c$ $(\leq c')$ is most powerful of all similar regions against $\sigma_1 > \sigma_0$ $(\sigma_1 < \sigma_0)$ was shown by Neyman and Pearson [1]

We consider first the case $\sigma_1 < \sigma_0$, and apply theorem 1 with λ a stepfunction having a single jump at ξ_1, that is,

$$\lambda(\xi) = \begin{matrix} 0 & \text{if} & \xi < \xi_1; \\ 1 & \text{if} & \xi \geq \xi_1. \end{matrix} \tag{21}$$

The region w given by (8) thus becomes

$$\frac{\exp\left[-\frac{1}{2\sigma_1^2}\Sigma(x_i - \xi_1)^2\right]}{\exp\left[-\frac{1}{2\sigma_0^2}\Sigma(x_i - \xi_1)^2\right]} \geq k', \tag{22}$$

which is equivalent to

$$\Sigma(x_i - \xi_1)^2 \leq k, \tag{23}$$

since $\sigma_1 < \sigma_0$. The size of the region (23), that is, its probability under the null hypothesis is a function of ξ and clearly attains its maximum when $\xi = \xi_1$. Thus all conditions of theorem 1 are satisfied provided we choose k so that the maximum size of (23) equals ϵ.

Before considering the case $\sigma_1 > \sigma_0$ we state for later reference the following:

LEMMA 1. *If $\sigma_1 > \sigma_0$ there exists an absolutely continuous non-decreasing function λ of bounded variation such that*

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma_0^2}(t - \xi)^2\right] d\lambda(\xi) = C \exp\left[-\frac{1}{2\sigma_1^2}(t - \xi_1)^2\right]. \tag{24}$$

This follows immediately from the well known representation of $\exp\left(-\frac{1}{a^2}t^2\right)$ as a Laplace transform by applying a translation, and is easily verified directly by substituting

$$\lambda'(\xi) = \exp\left[-\frac{1}{2(\sigma_1^2 - \sigma_0^2)}(\xi - \xi_1)^2\right]. \tag{25}$$

Now let $\sigma_1 > \sigma_0$ and $n > 1$. The region w given by (8) can be expressed in the form

$$\frac{\exp\left[-\frac{1}{2\sigma_1^2}\Sigma(x_i - \bar{x})^2\right]}{\exp\left[-\frac{1}{2\sigma_0^2}\Sigma(x_i - \bar{x})^2\right]} \cdot \frac{\exp\left[-\frac{n}{2\sigma_1^2}(\bar{x} - \xi_1)^2\right]}{\int \exp\left[-\frac{n}{2\sigma_0^2}(\bar{x} - \xi)^2\right] d\lambda(\xi)} \geq k'. \tag{26}$$

By lemma 1 there exists an absolutely continuous function λ for which the second factor is constant. For this λ (26) is equivalent to

$$\Sigma(x_i - \bar{x})^2 \geq c, \tag{27}$$

and since this is a similar region, the conditions of theorem 1 are satisfied provided c is chosen so as to give the correct level of significance.

We next consider the problem in which the random variables X_i $(i = 1, \cdots, n)$ are independently normally distributed with unknown means ξ_i and unknown variances σ_i^2. We wish to test the hypothesis $H_0 : \sigma_i = \sigma_{i0}$ for $i = 1, \cdots, n$ against the alternative $H_1 . \sigma_i = \sigma_{i1}$, $\xi_i = \xi_{i1}$. Feller [12] showed that there exist no similar regions for this problem. However, as we shall show now, when the critical regions are not required to be similar, non-trivial tests against H_1 do exist provided $\sigma_{i1} < \sigma_{i0}$ for at least one value of i.

Let us assume without loss of generality that $\sigma_{i1} < \sigma_{i0}$ for $i = 1, \cdots, m$; $\sigma_{i1} > \sigma_{i0}$ for $i = m + 1, \cdots, n$ where $n - m$ may be zero but where for the moment we shall assume $m > 0$ With $\lambda(\xi_1, \cdots, \xi_n) = \prod_{i=1}^{n} \lambda_i(\xi_i)$, the region (8) becomes

$$\prod_{i=1}^{m} \frac{\exp\left[-\frac{1}{2\sigma_{i1}^2}(x_i - \xi_{i1})^2\right]}{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma_{i0}^2}(x_i - \xi_i)^2\right] d\lambda_i(\xi_i)} \cdot \prod_{j=m+1}^{n} \frac{\exp\left[-\frac{1}{2\sigma_{j1}^2}(x_j - \xi_{j1})^2\right]}{\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma_{j0}^2}(x_j - \xi_j)^2\right] d\lambda_j(\xi_j)} \geq k. \tag{28}$$

For $\lambda_i(i = 1, \cdots, m)$ we take step functions with a single jump at ξ_{i1}, while for the remaining λ's we choose the absolutely continuous functions which make

the second factor constant and whose existence is guaranteed by lemma 1. The region (28) thus reduces to

$$\sum_{i=1}^{m}\left(\frac{1}{\sigma_{i1}^{2}}-\frac{1}{\sigma_{i0}^{2}}\right)(x_i-\xi_{i1})^2 \leq c. \tag{29}$$

Since the probability of the region (29) is independent of $\xi_{m+1}, \cdots, \xi_n$ and with varying $\xi_1, \cdots, \xi_m$ takes on its maximum when $\xi_i = \xi_{i1}$ it follows from theorem 1 that this region is most powerful for testing H_0 against H_1.

We still have to consider the case $m = 0$, that is, the case in which $\sigma_{i1} > \sigma_{i0}$ for all i. To treat this problem we adjoin to the variables $X_1, \cdots, X_n$ a random variable Y uniformly distributed between 0 and 1, that is, essentially a table of random numbers. In the space of $n + 1$ random variables we determine a region w according to (8), letting $\lambda(\xi_1, \cdots, \xi_n) = \prod_{i=1}^{n} \lambda_i(\xi_i)$ and choosing the λ's so as to make the left hand side of (8) equal to the right hand side. This is possible by lemma 1 and with this choice of the λ's the inequalities (9) become

$$\begin{aligned} k &\geq k \text{ in } w; \\ k &\leq k \text{ in } R_{n+1} - w, \end{aligned} \tag{30}$$

and hence they impose no restrictions on w. Thus any similar region of the correct size will satisfy the conditions of theorem 1. It follows that the region

$$w: 0 \leq y \leq \epsilon, \tag{31}$$

being a similar region of size ϵ, is most powerful. This result means that we do not use the observations $x_1, \cdots, x_n$ at all but consult a table of random numbers.

The situation just described occurs in other problems to which the same method of proof can be applied. It is therefore convenient for later reference to formulate the following

THEOREM 2. *Let H_0 be the hypothesis that the random variable X is distributed according to a probability density function f_θ with θ in ω, and let H_1 denote the alternative that X is distributed according to the density function g. Let Y be a random variable known to be uniformly distributed over the interval* [0, 1]. *If there exists a real valued function λ satisfying* (7) *for which*

$$g(x) = k \int_{\omega} f_\theta(x)\, d\lambda(\theta), \tag{32}$$

then the critical region $0 \leq y \leq \epsilon$ is most powerful for testing H_0 against H_1 at level of significance ϵ.

4. Testing equality of variances and the value of the circular serial correlation coefficient. For each $i = 1, \cdots, m$ let $X_{ij}(j = 1, \cdots, n_i)$ be a sample from a normal distribution with $E(X_{ij}) = \xi_i$ and $E(X_{ij} - \xi_i)^2 = \sigma_i^2$. We are con-

cerned with the hypothesis H_0 that $\sigma_1 = \sigma_2 = \cdots = \sigma_m$, where first we shall assume the ξ's to be known, so that without loss of generality we may assume them equal to 0. The alternative hypothesis specifies $\sigma_i = \sigma_{i1}$, $i = 1 \cdots m$. Let σ^2 denote the unknown common variance under H_0 and let $\lambda(\sigma)$ be a step function with a single jump at a point σ_0 to be determined later With $k = \prod_{i=1}^{m} \left(\frac{\sigma_0}{\sigma_{i1}}\right)^{n_i}$, the test (8) takes on the form

$$\frac{\exp\left[-\frac{1}{2}\sum_{i,j}\frac{x_{ij}^2}{\sigma_{i1}^2}\right]}{\exp\left[-\frac{1}{2\sigma_0^2}\sum_{i,j}x_{ij}^2\right]} \geq 1, \tag{33}$$

or equivalently

$$\frac{\sum_{i,j} x_{ij}^2}{\sum_{i,j}\frac{1}{\sigma_{i1}^2}x_{ij}^2} \geq \sigma_0^2 \tag{34}$$

Since the function on the left hand side is homogeneous of degree 0 in the x's, this is a similar region and the conditions of theorem 1 are therefore satisfied provided the region has the correct size. This can be achieved for any level of significance ϵ by proper choice of σ_0^2.

As stated earlier, the conditions of theorem 1 imply that the size of the critical region is equal to ϵ at all points of increase of λ. As a consequence, if the size equals ϵ at only a finite number of points of ω, λ must be a step function. Also if each point of a certain interval is a point of increase of λ, the critical region must be similar over that interval (and, if the functions involved are analytic, the region must be similar over ω). However, the last problem shows that the converse of neither of these two statements is correct. For the region (34) is a similar region although the corresponding λ has only a single point of increase.

Next we consider the hypothesis of equality of variances without assuming the means to be known For the case $m = 2$ the most powerful similar region was obtained by Neyman and Pearson [1] We assume first that $n_i > 1$ for all i, and we take $\lambda(\sigma, \xi_1, \cdots, \xi_m) = \lambda_0(\sigma)\prod_{i=1}^{m}\lambda_i(\xi_i)$, with $\lambda_0(\sigma)$ as before a step function with a single jump at a point σ_0 to be determined later Suppose now that $\sigma_0 > \sigma_{i1}$ for $i = 1, \cdots, s$; $\sigma_0 < \sigma_{i1}$ for $i = s + 1, \cdots, m$, $\sigma_{11} \leq \sigma_{21} \leq \cdots$ where $0 \leq s \leq m$ and s depends on σ_0. Then define

$$\lambda_i(\xi_i) = \begin{cases} 0 & \text{if } \xi_i < \xi_{i1}; \\ 1 & \text{if } \xi_i \geq \xi_{i1} \end{cases} \tag{35}$$

for $i = 1, \cdots, s$ and use lemma 1 for $i = s + 1, \cdots, m$

For proper choice of k the critical region will then be determined by the inequality

$$(36)\qquad \sum_{i=s+1}^{m}\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_{i1}^2}\right)\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2-\sum_{i=1}^{s}\left(\frac{1}{\sigma_{i1}^2}-\frac{1}{\sigma_0^2}\right)\sum_{j=1}^{n_i}(x_{ij}-\xi_{i1})^2\geq 0$$

The probability of this region computed under H_0, is independent of $\xi_{s+1}, \cdots, \xi_n$ and for any σ attains its maximum when $\xi_i = \xi_{i1}$ $(i = 1, \cdots, s)$. Since the probability of the region is independent of σ when $\xi_i = \xi_{i1}$ for $i = 1, \cdots, s$, the conditions of theorem 1 are again established That for $\xi_i = \xi_{i1}$ the size of (36) goes continuously from 0 to 1 with decreasing σ_0 is easily checked since at the only doubtful points $\sigma_0 = \sigma_{i1}$ (where the value of s changes), the corresponding coefficient $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_{i1}^2}$ passes through 0.

We still have to consider the case that some of the n_i are equal to 1 If $n_i = 1$ for some $i \leq s$ there is no change whatever, while if $n_i = 1$ for some $i > s$, the corresponding term in (36) vanishes It follows easily that if $n_i > 1$ for at least one value of $i > 1$ the solution (36) is valid. On the other hand, if $n_i = 1$ for all $i > 1$, we can apply theorem 2 by taking $\sigma_0 = \sigma_{i1}$, $\lambda_1(\xi_1)$ as a step function with a single jump at ξ_{11} and the remaining $\lambda_i(\xi_i)$ according to lemma 1. It thus follows that for this problem no non-trivial test exists.

The following problem can be reduced to the hypothesis of equality of variances with means assumed known: Under the null hypothesis $X_1, \cdots, X_n$ have a joint multivariate normal distribution with density $C \exp\left[-\frac{1}{2\sigma^2}\Sigma\, a_{ij}x_ix_j\right]$ where the a's are known and where σ is an unknown scale factor. Under H_1 the X's have a joint multivariate normal distribution with density C' exp $\left[-\frac{1}{2}\Sigma b_{ij}x_ix_j\right]$ A number of hypotheses specifying the value of one or several correlation coefficients have this form. The most powerful test of H_0 against H_1 is given by

$$(36)\qquad \frac{\Sigma b_{ij}x_ix_j}{\Sigma a_{ij}x_ix_j}$$

as is easily shown by applying a non-singular linear transformation which reduces $\Sigma b_{ij}x_ix_j$ to diagonal form and $\Sigma a_{ij}x_ix_j$ to a sum of squares, or by applying directly the method of proof of the earlier problem.

A corresponding reduction when the X's have a common but unknown mean is usually impossible. One problem of this kind for which the solution is simple is the hypothesis specifying the value of a serial correlation coefficient in a circular population. The most powerful similar region for testing this hypothesis was obtained in [7] Consider the probability density function

$$(37)\qquad C\exp\left[-\alpha\left\{\sum_{i=1}^{n}(x_i-\xi)-\delta(x_{i+1}-\xi)\right\}^2\right],$$

$$(x_{n+1} = x_1), \qquad |\delta| < 1,$$

and let H_0 specify $\delta = \delta_0$ while H_1 assigns to the parameters the values α_1, ξ_1, δ_1. Then the most powerful test of H_0 against H_1 is

$$(38) \qquad \begin{aligned} \frac{\Sigma(x_i - \bar{x})(x_{i+1} - \bar{x})}{\Sigma(x_i - \bar{x})^2} &\geq k \quad \text{if} \quad \delta_1 > \delta_0, \\ \frac{\Sigma(x_i - \xi_1)(x_{i+1} - \xi_1)}{\Sigma(x_i - \xi_1)^2} &\leq k' \quad \text{if} \quad \delta_1 < \delta_0 \end{aligned}$$

We shall omit the proof of this result, since the method is the same as in the other problems considered in this section.

5. Student's hypothesis and some generalizations. As the principal result of the present section we shall prove that for testing Student's hypothesis against a simple alternative the most powerful test is a non-similar region of the form

$$(39) \qquad \Sigma(X_i - \eta)^2 \leq k,$$

if the level of significance ϵ is less than or equal to $\frac{1}{2}$. Here η and k depend on ϵ and on the alternative, and they will not be determined explicitly. It will be shown also that if ϵ is greater than or equal to $\frac{1}{2}$, Student's test is most powerful. These results will be extended rather easily to the general univariate linear hypothesis. The corresponding investigation for similar regions was carried through for Student's hypothesis by Neyman and Pearson [1] while the extension to a general linear hypothesis is contained in a paper by Hsu [13].

The proof of the main result mentioned above is rather lengthy. We shall begin by proving two lemmas.

LEMMA 2. *Let $Y_1, \cdots, Y_n$ be n independent random variables, normally distributed with* 0 *mean and unit variance, and let*

$$(40) \qquad \begin{aligned} P(a, k) &= P\left\{\sum_{i=1}^{n} (Y_i - a)^2 \leq (n - k)a^2\right\}; \\ \varphi(k) &= \sup_a P(a, k) \quad \textit{for} \quad 0 < k < n, \quad 0 < a \end{aligned}$$

Then for each k there exists $a(k)$ such that

$$(41) \qquad P(a(k), k) = \varphi(k).$$

PROOF. If $Z_i = Y_i/a$, $(i = 1, \cdots, n)$ the Z's are independently normally distributed with zero mean and variance $1/a^2$ and (40) may be written as

$$(42) \qquad P(a, k) = P\{\Sigma(Z_i - 1)^2 \leq n - k\}$$

Hence it is seen that for any k, $P(a, k)$ tends to zero as a tends to either zero or infinity. This proves the lemma since for any k, $P(a, k)$ is a continuous function of a.

LEMMA 3. *Given any ϵ, $0 < \epsilon < \frac{1}{2}$ there exists $k(\epsilon)$ between zero and n such that $\varphi(k(\epsilon)) = \epsilon$.*

PROOF. The proof will be given in a number of steps

(i) $\varphi(k) \to \frac{1}{2}$ as $k \to 0$.

Clearly $P(a, k)$ never exceeds $\frac{1}{2}$. The result will therefore follow if we exhibit a sequence a_k such that $P(a_k, k) \to \frac{1}{2}$ as $k \to 0$. Let $a_k = 1/\sqrt{k}$ Then

$$P(a_k, k) = P\{\sqrt{k}\, \Sigma Y_i^2 - 2\Sigma Y_i + \sqrt{k} \leq 0\}. \tag{43}$$

The right hand side is a continuous function of k and therefore tends to

$$P\{\Sigma Y_i \geq 0\} = \tfrac{1}{2}, \tag{44}$$

as k tends to zero.

(ii) $\varphi(k) \to 0$ as $k \to n$.

Consider $P(a, k)$ as in (42). Written as an integral of the probability density of the Z's, the region of integration is independent of a and its volume tends to 0 as k tends to n. On the other hand the probability density depends on a but is uniformly bounded over the region of integration if $k > 0$, and hence the result follows.

(iii) If $0 < k_0$, $P(a, k)$ tends to zero uniformly for k in the interval $k_0 \leq k \leq n$ as a tends to zero or infinity.

This follows from the fact that $0 \leq P(a, k) \leq P(a, k_0)$ since $P(a, k_0)$ tends to 0 as a tends to zero or infinity.

(iv) Given k_0 and k_1 there exist numbers a_0 and a_1 with $0 < a_0 < a_1 < \infty$ such that $0 < k_0 \leq k \leq k_1 < n$ implies $a_0 \leq a(k) \leq a_1$.

If this were not true there would exist a sequence $k^{(i)}$ with $k_0 \leq k^{(i)} \leq k_1$ and $a(k^{(i)})$ tending to infinity or zero. Then $\varphi(a(k^{(i)}))$ would tend to zero by (iii). On the other hand consider $P(1, k)$ for $k_0 \leq k \leq k_1$. This is a continuous non-vanishing function of k and hence attains its lower bound m for some k in $k_0 \leq k \leq k_1$. Therefore m is positive and we have a contradiction.

(v) Given any k_0, k_1 with $0 < k_0 < k_1 < n$, $\varphi(k)$ is continuous on the interval $[k_0, k_1]$.

To see this, select a_0 and a_1 in accordance with (iv). Then $P(a, k)$ is uniformly continuous in the rectangle $a_0 \leq a \leq a_1$, $k_0 \leq k \leq k_1$. Given $\eta > 0$ let δ be such that $|k' - k''| < \delta$ implies $|P(a, k') - P(a, k'')| < \eta$. Then $\varphi(k') \geq P(a(k''), k') \geq P(a(k''), k'') - \eta = \varphi(k'') - \eta$, and by symmetry $\varphi(k'') \geq \varphi(k') - \eta$, which establishes the continuity of φ.

The proof of the lemma is now immediate. For let $0 < \epsilon < \frac{1}{2}$. It follows from (i) and (ii) that there exist k_0 and k_1 such that

$$\varphi(k_0) \leq \epsilon/2, \qquad \varphi(k_1) \geq \epsilon + \tfrac{1}{2}(\tfrac{1}{2} - \epsilon),$$

and hence by (v) there exists $k(\epsilon)$ for which $\varphi(k(\epsilon)) = \epsilon$.

Let us now consider Student's hypothesis. The random variables $X_1, \cdots, X_n$ are a sample from a normal distribution which under H_0 has mean 0 and unknown variance σ^2, while under H_1 the mean is ξ_1 and the variance σ_1^2. Without loss of generality we shall assume $\xi_1 > 0$. Applying theorem 1 with λ a step-function having a single jump at a point $\sigma_0 > \sigma_1$ to be determined later, we obtain the critical region in the form

$$(45) \qquad \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) \Sigma X_i^2 - 2 \frac{\xi_1}{\sigma_1^2} \Sigma X_i \leq c$$

Let $Y_i = X_i/\sigma$ so that under H_0 the Y's are distributed with zero mean and unit variance. Then (45) becomes

$$(46) \qquad \Sigma Y_i^2 - 2 \frac{\xi_1}{\sigma(1 - \sigma_1^2/\sigma_0^2)} \Sigma Y_i \leq \frac{c}{\sigma^2},$$

which may be written as

$$(47) \qquad \Sigma(Y_i - a)^2 \leq (n - k)a^2,$$

where

$$(48) \qquad a = \frac{\xi_1}{\sigma(1 - \sigma_1^2/\sigma_0^2)}, \qquad k = \frac{-c}{\xi_1^2}\left(1 - \frac{\sigma_1^2}{\sigma_0^2}\right)^2.$$

As σ varies from 0 to ∞, a goes from ∞ to 0. Let $P(a, k)$, $\varphi(k)$ and $a(k)$ be defined as in lemma 2. Given the level of significance ϵ $(0 < \epsilon < \frac{1}{2})$, let k^* and a^* be determined according to lemma 2 and 3 so that

$$(49) \qquad \varphi(k^*) = \epsilon \text{ and } P(a^*, k^*) = \varphi(k^*).$$

We now select $\sigma_0 > \sigma_1$ and c so that

$$(50) \qquad a^* = \frac{\xi_1}{(1 - \sigma_1^2/\sigma_0^2)\sigma_0} \quad \text{and} \quad k^* = \frac{c}{\xi_1^2}\left(1 - \frac{\sigma_1^2}{\sigma_0^2}\right).$$

We have to show that for this choice of σ_0 and c the size of the critical region attains its maximum when $\sigma = \sigma_0$ and that this maximum size is ϵ. Substituting from (50) we express the region (47) in the form

$$(51) \qquad \Sigma\left(Y_i - \frac{\sigma_0}{\sigma} a^*\right)^2 \leq (n - k^*) \frac{\sigma_0^2}{\sigma^2} a^{*2}.$$

Thus the probability of the region is

$$(52) \qquad P\left(\frac{\sigma_0}{\sigma} a^*, k^*\right).$$

As σ varies, (52) attains its maximum when $\frac{\sigma_0}{\sigma}a^* = a(k^*) = a^*$, that is, when $\sigma = \sigma_0$ and the maximum value of (52) is $\varphi(k^*) = \epsilon$.

This derivation is valid even when $n = 1$, i.e., when the hypothesis $\xi = 0$ is to be tested by observing only a single random variable X, known to be normally distributed but whose mean ξ and variance are unknown. For this problem no similar region exists. However, critical regions of the form $0 < \xi_1 - a < x < \xi_1 + b$ will give any level of significance $< \frac{1}{2}$ for proper choice of a and b, while the power of such regions will tend to 1 as σ_1 tends to 0. Therefore, the power of the most powerful test will be close to 1 if σ_1 is sufficiently small.

Having completed the discussion of the case $\epsilon < \frac{1}{2}$ let us next suppose that $\epsilon \geq \frac{1}{2}$. We shall need the following

LEMMA 4. *Let c and α_1 be positive constants. Then there exists a function f such that $f(\alpha) = 0$ when $\alpha < \alpha_1$ and such that for all $w > 0$*

$$\int_0^\infty e^{-\alpha w} f(\alpha)\, d\alpha = k e^{-\alpha_1 w - c\sqrt{w}}. \tag{53}$$

This follows from the well known representation of $e^{-c\sqrt{w}}$ as a Laplace transform by applying a translation. (53) can be checked directly by substituting

$$f(\alpha) = \frac{c e^{-(c^2/4(\alpha-\alpha_1))}}{(\alpha - \alpha_1)^{3/2}} \quad \text{for} \quad \alpha \geq \alpha_1. \tag{54}$$

Applying theorem 1 to Student's hypothesis, where again we shall assume ξ_1 to be positive, for proper choice of k we obtain from (9)

$$\frac{\exp\left[-\frac{1}{2\sigma_1^2}\Sigma X_i^2 + \frac{\xi_1}{\sigma_1^2}\Sigma X_i\right]}{\int_0^\infty \exp\left[-\frac{1}{2\sigma^2}\Sigma X_i^2\right]\frac{1}{\sigma^n}\, d\lambda(\sigma)} \geq 1. \tag{55}$$

It follows from lemma 4 that for any positive c there exists a non-decreasing function λ of bounded variation with $\lambda(\sigma)$ constant for $\sigma > \sigma_1$, such that

$$\int_0^\infty \exp\left[-\frac{1}{2\sigma^2}\Sigma X_i^2\right]\frac{1}{\sigma^n}\, d\lambda(\sigma) = \exp\left[-\frac{1}{2\sigma_1^2}\Sigma X_i^2 - c\sqrt{\Sigma x_i^2}\right]. \tag{56}$$

For this choice of λ, (55) reduces to

$$\exp\left[\frac{\xi_1}{\sigma_1^2}\Sigma x_i\right] \geq \exp\left[-c\sqrt{\Sigma x_i^2}\right], \tag{57}$$

and hence to

$$\frac{\Sigma x_i}{\sqrt{\Sigma x_i^2}} \geq c' \tag{58}$$

This is a similar region and therefore most powerful for testing Student's hypothesis against H_1. By adjusting c, the size of the region can be made equal to any $\epsilon \geq \frac{1}{2}$.

The argument for $\epsilon > \frac{1}{2}$ must be modified slightly in the case $n = 1$, that is, when we want to test Student's hypothesis on the basis of a single observation. Let us adjoin to the variable X a random variable Y known to be uniformly distributed over the interval [0, 1]. Using the same λ and k as before, (58) becomes

$$\frac{x}{|x|} \geq c' \tag{59}$$

For $c' = -1$ the critical region includes all points (x, y) for which x is positive while (59) places no restriction on which of the remaining points to include in the critical region The similar region

$$x \geq 0, \quad x < 0, \quad 0 < y < 2(\epsilon - \tfrac{1}{2}) \tag{60}$$

therefore satisfies all conditions of theorem 1 and hence is most powerful

In extending these results to the general linear hypothesis, we shall assume the hypothesis reduced to canonical form [14, 15]. We shall therefore assume that $X_1, \cdots, X_n$ are normally distributed with common variance which is unknown under H_0 and has the value σ_1^2 under H_1. Furthermore, under H_0, $E(X_i) = 0$ for $i = 1, \cdots, s, s + 1, \cdots, m$, $E(X_i)$ unknown for $i = m + 1, \cdots, n$ while under H_1 $E(X_i) = 0$ for $i = s + 1, \cdots, m$, $E(X_i) = \xi_{i1}$ for the remaining values of i.

For $\epsilon < \frac{1}{2}$ we shall consider critical regions of the form

$$\frac{\exp\left\{-\dfrac{1}{2\sigma_1^2}\left[\sum_{i=1}^{s}(x_i - \xi_{i1})^2 + \sum_{i=s+1}^{m} x_i^2 + \sum_{i=m+1}^{n}(x_i - \xi_{i1})^2\right]\right\}}{\exp\left\{-\dfrac{1}{2\sigma_1^2}\left[\sum_{i=1}^{s} x_i^2 + \sum_{i=s+1}^{m} x_i^2 + \sum_{i=m+1}^{n}(x_i - \xi_{i1})^2\right]\right\}} \geq k, \tag{61}$$

which are obtained from (8) by substituting for λ a step-function with a single jump at the parameter point $(\sigma_i, \xi_{m+1,1}, \cdots, \xi_{n,1})$. Making an orthonormal transformation from $x_s, \cdots, x_s$ to $y_1, \cdots, y_s$ such that $y_1 = \dfrac{\sum_{i=1}^{s} \zeta_{i1}\lambda_i}{\sqrt{\Sigma\zeta_{i1}^2}}$ and letting $y_i = x_i$ for $i = s + 1, \cdots, m$, $y_i = x_i - \xi_{i1}$ for $i = m + 1, \cdots, n$, (61) reduces to

$$\frac{\exp\left\{-\dfrac{1}{2\sigma_1^2}\left[\sum_{i=1}^{n} y_i^2 - 2y_1\sqrt{\sum_{i=1}^{s}\xi_{i1}^2}\right]\right\}}{\exp\left\{-\dfrac{1}{2\sigma_0^2}\sum_{i=1}^{n} y_i^2\right\}} \geq c. \tag{62}$$

For $\sigma_0 > \sigma_1$ we can rewrite (62) as

$$\sum_{i=m+1}^{n} y_i^2 \leq \frac{1}{\sigma_1^2 - \sigma_0^2}\left[c + \frac{2y_1\sqrt{\Sigma\xi_{i1}^2}}{\sigma_1^2}\right] - \sum_{i=1}^{m} y_i^2, \tag{63}$$

and we see that under H_0 for any σ the size of this region considered as a function of the unknown means of $Y_{m+1}, \cdots, Y_n$ takes on its maximum when these means are zero, i e. when $\xi_i = \xi_{i1}$ for $i = m + 1, \cdots, n$. For these maximizing values of the means the existence of a suitable σ_0 and c follows from the corresponding result in connection with Student's hypothesis

Thus the most powerful test for testing H_0 against H_1 at level of significance $\epsilon = \frac{1}{2}$ has the form

$$\sum_{i=1}^{s}\left[x_i - \frac{\xi_{i1}}{1 - \sigma_1^2/\sigma_0^2}\right]^2 + \sum_{i=s+1}^{m} x_i^2 + \sum_{i=m+1}^{n}(x_i - \xi_{i1})^2 \leq c. \tag{64}$$

It is interesting that the variables $X_i (i = m + 1, \cdots, n)$ which may be discarded when considerations are restricted to similar regions [18], do contribute to the power when similarity is not required. The same phenomenon also occurs in certain problems considered earlier in this paper.

For the case $\epsilon \geq \frac{1}{2}$, let us take

$$\lambda(\sigma, \xi_{m+1}, \cdots, \xi_n) = \lambda(\sigma) \prod_{i=m+1}^{n} \lambda_i(\xi_i \mid \sigma). \tag{65}$$

We shall select $\lambda(\sigma)$ such that $\lambda(\sigma)$ is constant when $\sigma \geq \sigma_1$. Hence it is enough to define $\lambda_i(\xi_i \mid \sigma)$ for $\sigma < \sigma_1$. For any $\sigma < \sigma_1$ there exists by lemma 1 a function $\lambda_i(\xi_i \mid \sigma)$ such that

$$\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}(x_i - \xi_i)^2\right] d\lambda_i(\xi_i \mid \sigma) = k \exp\left\{-\frac{1}{2\sigma_1^2}(x_i - \xi_{i1})^2\right\} \tag{66}$$

For this choice of the λ_i, (9) becomes

$$\frac{\exp\left\{-\frac{1}{2\sigma_1^2}\left[\sum_{i=1}^{s}(x_i - \xi_{i1})^2 + \sum_{i=s+1}^{m} x_i^2\right]\right\}}{\int_0^{\infty} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{m} x_i^2\right\} d\lambda(\sigma)} \geq k'. \tag{67}$$

Next we chose $\lambda(\sigma)$ according to lemma 4 such that

$$\int_0^{\infty} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{m} x_i^2\right]\frac{1}{\sigma^m}\, d\lambda(\sigma) = \exp\left[-\frac{1}{2\sigma_1^2}\sum_{i=1}^{m} x_i^2 - c\sqrt{\sum_{i=1}^{m} x_i^2}\right], \tag{68}$$

thus, by proper choice of k', reducing (67) to

$$\frac{\sum_{i=1}^{s} \xi_{i1} x_i}{\sqrt{\sum_{i=1}^{m} x_i^2}} \geq -c. \tag{69}$$

The probability of this region under H_0 is independent of $\xi_{m+1}, \cdots, \xi_n$ and σ, and hence (69) is most powerful for testing H_0 against H_1

Let us return once more to the problem of testing Student's hypothesis against a simple alternative $\xi = \xi_1$, $\sigma = 1$ and let us assume as known that $\sigma \leq 1$. No use can be made of this knowledge if consideration is restricted to similar regions. For the probability of first kind error is an analytic function of σ, and consequently, if a test is similar with respect to all values of σ which are ≤ 1, it is similar with respect to all values of σ. Let us now consider this problem without the restriction of similarity If $\epsilon \geq \frac{1}{2}$, the knowledge concerning σ does not enable us to find a test which is more powerful than that given by (58), since the function $\lambda(\sigma)$ on which (58) was based had all its points of increase for $\sigma \leq 1$.

On the other hand we may expect improvement for $\epsilon < \frac{1}{2}$ since the most powerful test in this case was based on a function λ with a single point of increase

$\sigma_0 > 1$ which is no longer admitted as a possible value of σ. If, instead, we take for λ the step function with a single jump at $\sigma = 1$ we obtain the critical region

$$\frac{\exp\left[-\frac{1}{2}\sum(x_i - \xi_1)^2\right]}{\exp\left[-\frac{1}{2}\sum x_i^2\right]}, \tag{70}$$

which is equivalent to

$$\bar{x} \geq c. \tag{71}$$

Here $c > 0$ since $\epsilon < \frac{1}{2}$, and therefore, when $\xi = 0$ the probability of (71) is an increasing function of σ and hence takes on its maximum at $\sigma = 1$ It follows from theorem 1 that (71) is most powerful under the conditions stated.

In the opposite problem in which it is known that $\sigma \geq 1$, the situation is reversed. For $\epsilon \leq \frac{1}{2}$ no improvement over (45) is possible while for $\epsilon > \frac{1}{2}$ we can use for λ the step function with a single step at $\sigma = 1$ thus obtaining the critical region (70) but this time with $c < 0$. When $\xi = 0$ the probability of this region is a decreasing function of σ and it follows that (70) is most powerful in this case.

Similar remarks apply to other problems We mention as one further example a modification of the Behrens-Fisher problem Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ be independently normally distributed, the X's with mean ξ and variance σ^2, the Y's with mean η and variance τ^2, all four parameters being unknown. We wish to test, at level of significance $\epsilon \leq \frac{1}{2}$, the hypothesis $\xi = \eta$ against the simple alternative $\xi = \xi_1$, $\eta = \eta_1$, $\sigma = 1$, $\tau = 1$, where $\xi_1 \neq \eta_1$ and we assume it known that $\sigma \leq 1$, $\tau \leq 1$. Basing the test on a step function λ with a single jump at $\sigma = 1$, $\tau = 1$, $\xi = \dfrac{n\xi_1 + m\eta_1}{m + n}$ we obtain for w the region

$$\frac{\exp\left[-\frac{1}{2}\sum(x_i - \xi_1)^2 - \frac{1}{2}\sum(y_i - \eta_1)^2\right]}{\exp\left[-\frac{1}{2}\sum\left(x_i - \frac{n\xi_1 + m\eta_1}{n+m}\right)^2 - \frac{1}{2}\sum\left(y_i - \frac{n\xi_1 + m\eta_1}{n+m}\right)^2\right]} \geq k, \tag{73}$$

which is equivalent to

$$\bar{y} - \bar{x} \geq c \qquad (c > 0), \tag{74}$$

if we assume, as we may without loss of generality, that $\eta_1 > \xi_1$. When $\eta = \xi_2$, $\bar{Y} - \bar{X}$ is normally distributed with zero mean and variance $\dfrac{\sigma^2}{n} + \dfrac{\tau^2}{m}$. Therefore the probability of (74) is an increasing function of $\dfrac{\sigma^2}{n} + \dfrac{\tau^2}{m}$ and hence attains its maximum when $\sigma = \tau = 1$. It follows from theorem 1 that the region (74) is most powerful for the problem under consideration.

6. Admissibility. The general problem to be considered in this paper has been formulated in section 1: To obtain a region w

$$\text{maximizing} \qquad \int_w g(x)\,dx \tag{75}$$

subject to the restriction

$$\int_w f_\theta(x)\,dx \leq \epsilon \qquad \textit{for all } \theta \in \omega. \tag{76}$$

Since for any particular such problem there may exist several essentially different regions satisfying these conditions, it may happen that there exists a region w' such that

$$\int_{w'} g(x)\,dx = \int_w g(x)\,dx, \tag{77}$$

and

$$\int_{w'} f_\theta(x)\,dx \leq \int_w f_\theta(x)\,dx \qquad \textit{for all } \theta \in \omega, \tag{78}$$

with inequality holding for some θ. Clearly w' is preferable to w. In this case, following the definition of Wald [4], we say that w is not admissible. We shall rule out this possibility for a large class of problems by proving

THEOREM 3. *If w satisfies the conditions of theorem* 1, *and if the set of points x for which equality holds in* (8) *has measure zero, then any region satisfying* (75) *and* (76) *differs from w only on a set of measure zero.*

PROOF. Without loss of generality we shall assume λ of theorem 1 to be a distribution function. Then

$$h(x) = \int_\omega f_\theta(x)\,d\lambda(\theta)$$

is a completely specified probability density function, and w is the unique[3]—up to a set of measure zero—most powerful test for testing the simple hypothesis $H_0'\!:\!h$ against the simple alternative $H_1\!:\!g$. Suppose now that w' satisfies (75) and (76). Then

$$\int_{w'} h(x)\,dx \leq \epsilon, \tag{79}$$

and w' is most powerful for testing H_0' against H_1. It follows that w' differs from w at most by a null set.

Earlier we enlarged the problem of testing by adjoining to the original random variable X a random variable with a known distribution. This is equivalent to the following modification of the original problem Instead of defining a test to be a critical region (of rejection) in the space of x, we define it to be a critical

[3] One sees this easily from Neyman and Pearson's proof of the fundamental lemma [1], by using the assumption that the set of points for which equality holds in (8), has measure zero.

function φ $(0 \leq \varphi(x) \leq 1)$ which with every point x associates a probability of rejection $\varphi(x)$. If x is observed, the hypothesis is rejected with probability $\varphi(x)$ according to a table of random numbers In the case where random numbers are not employed, φ merely becomes the characteristic function of the set w

We shall now state a theorem which will prove admissibility for all but one of those problems treated in sections two to five, to which theorem 3 does not apply.

THEOREM 4. *Suppose $\omega = \{\theta\}$ is a subset of an s-dimensional Euclidean space, and that for any measurable function φ and for any set S which has positive measure and is contained in ω*

$$\int \varphi(x) f_\theta(x)\, dx = c \qquad \text{for } \theta \in S \tag{80}$$

implies

$$\int \varphi(x) f_\theta(x)\, dx = c \qquad \text{for } \theta \in \omega. \tag{81}$$

(Here and in all that follows whenever a region of integration is not indicated, the integral extends over the whole x space). Suppose further that φ is a critical function satisfying the conditions of theorem 1 and that the set S_0 of points of increase of λ has positive measure Then φ is admissible.

PROOF. If φ were not admissible there would exist φ_1 with

$$\int \varphi_1(x)\, g(x)\, dx = \int \varphi(x) g(x)\, dx, \tag{82}$$

$$\int \varphi_1(x) f_\theta(x)\, dx \leq \int \varphi(x) f_\theta(x)\, dx \qquad \text{for all } \theta \in \omega; \tag{83}$$

$$\int \varphi_1(x) f_\theta(x)\, dx < \int \varphi(x) f_\theta(x)\, dx \qquad \text{for some } \theta \in \omega \tag{84}$$

The set T of points θ for which (84) holds, differs from ω at most by a null set. For

$$\int [\varphi_1(x) - \varphi(x)] f_\theta(x)\, dx = 0 \qquad \text{for } \theta \in \omega - T, \tag{85}$$

and if $\omega - T$ had positive measure, (85) would hold for all $\theta \in \omega$.

Let h and H_0' be defined as in the proof of theorem 3. Since S has positive measure, it follows that

$$\epsilon = \int \varphi(x) h(x)\, dx > \int \varphi_1(x) h(x)\, dx = \eta, \qquad \text{say.} \tag{86}$$

Let $\varphi_2(x) = \min \left[1, \varphi_1(x) + \epsilon - \eta \right]$. Then

$$\int \varphi_2(x) h(x)\, dx \leq \epsilon \tag{87}$$

and

$$\int \varphi_2(x)g(x)\,dx > \int \varphi_1(x)g(x)\,dx. \tag{88}$$

But φ_1 is most powerful for testing H_0' against H_1 and we have a contradiction.

By applying theorems 3 and 4 one can easily show for all but one of the problems treated in sections three to five that the tests obtained there are admissible. The one exception occurs when testing equality of variances. Simplifying the notation, since we are now concerned with a special case, we shall assume that $X_i(i = 1, \cdots, n)$, $Y_1, \cdots, Y_r$ are independently and normally distributed, the X's with mean ξ_0 and variance σ_0^2, Y_i with mean ξ_i and variance σ_i^2, all parameters being unknown. We wish to test the hypothesis of equality of variances against the simple alternative

$$H_1 : \xi_i = \xi_{i1}, \qquad \sigma_i = \sigma_{i1} \qquad (i = 0, \cdots, r),$$

with

$$\sigma_{01} < \sigma_{11} < \cdots < \sigma_{r1}.$$

We shall first consider the case $n = 1$, and prove admissibility of the critical function

$$\varphi(x, y_1, \cdots, y_r) = \epsilon \tag{89}$$

by using a different distribution function for the parameters from the one used earlier. With some specialization of the distribution function, (8) becomes for our problem

$$\frac{\exp\left\{-\dfrac{1}{2\sigma_{01}}(x-\xi_{01})^2 - \dfrac{1}{2}\sum_{i=1}^{r}\dfrac{1}{\sigma_{i1}^2}(y_i-\xi_{i1})^2\right\}}{\displaystyle\int\frac{1}{\sigma^{r+1}}\left\{\int\exp\left[-\frac{1}{2\sigma^2}(x-\xi_0)^2\,d\lambda_\sigma^{(0)}(\xi_0)\right]\cdot\prod_{i=1}^{r}\int\exp\left[-\frac{1}{2\sigma^2}(y_i-\xi_i)^2\right]d\lambda_\sigma^{(i)}(\xi_i)\right\}d\mu(\sigma)} \geq k \tag{90}$$

For any $\sigma < \sigma_{01}$ we select the $\lambda_\sigma^{(i)}(\xi_i)$ according to lemma 1. If we then take for μ the uniform distribution over $(\sigma_{01} - 1, \sigma_{01})$ the left hand side of (90) will reduce to k. Admissibility of the critical function (89) then follows from theorem 4.

That a constant critical function is not admissible in the case $n > 1$ is easily seen if one compares it for instance with the critical region

$$\left|\frac{\bar{x} - \xi_{01}}{\sqrt{\Sigma(x_1 - \bar{x})^2}}\right| \leq c. \tag{91}$$

We shall not obtain a complete family of admissible tests (cf. [4]) for the case $n > 1$ but we shall show that this problem is equivalent to the following one: To find a complete class of unbiased admissible tests for the hypothesis specifying

the mean and variance of a normal distribution on the basis of a sample from this distribution, the class of alternatives being the totality of univariate normal distributions.

Let $n > 1$ and let φ be any most powerful critical function for testing the hypothesis of equality of variances against H_1. If φ corresponds to the level of significance ϵ and if β_φ denotes the power of φ, we have

$$\beta_\varphi(\sigma, \sigma, \cdots, \sigma, \xi_0, \xi_1, \cdots, \xi_r) \leq \epsilon \tag{92}$$

for all admissible values of the arguments It also follows from section 4 that

$$\beta_\varphi(\sigma_{01}, \sigma_{11} \cdots, \sigma_{r1}, \xi_{01}, \xi_{11}, \cdots, \xi_{r1}) = \epsilon. \tag{93}$$

Consider for a moment the hypothesis $H_0': \sigma_i = \sigma_{01} (i = 0, \cdots, r)$, $\xi_0 = \xi_{01}$, ξ_i unspecified for $i = 1, \cdots, r$. It is easily seen that the maximum power for testing H_0' against H_1 is ϵ. Therefore any most powerful test for testing H_0 against H_1 is also most powerful for testing H_0' against H_1, and in particular this holds for φ. Furthermore, it follows easily from theorem 4 that for any most powerful test of H_0' against H_1 the probability of an error of the first kind must be identically equal to ϵ Therefore

$$\beta_\varphi(\sigma_{01}, \cdots, \sigma_{01}, \xi_{01}, \xi_1, \cdots, \xi_r) = \epsilon \text{ for all } \xi_1, \cdots, \xi_r \tag{94}$$

But (94) is equivalent to the condition that φ is similar with respect to $\xi_1, \cdots, \xi_r$, and it follows [12] that φ is a function of $x_1, \cdots, x_n$ only. The problem is therefore reduced to that of finding all admissible critical functions $\varphi(x_1, \cdots, x_n)$ satisfying

$$\beta_\varphi(\sigma_{01}, \xi_{01}) = \epsilon, \qquad \beta_\varphi(\sigma_0, \xi_0) \leq \epsilon \text{ for all } \sigma_0, \xi_0. \tag{95}$$

That this problem in turn is equivalent to the one stated above is immediate when one considers the complementary critical functions $1 - \varphi$

REFERENCES

[1] J NEYMAN AND E. S. PEARSON, "On the problem of the most efficient tests of statistical hypotheses," *Roy Soc Phil. Trans.*, Ser A, Vol 231 (1933), p 289

[2] A. WALD, "Test of statistical hypotheses concerning several parameters when the number of observations is large," *Am Math Soc. Trans*, Vol 54 (1943), p. 426

[3] A. WALD, "Statistical decision functions which minimize the maximum risk," *Annals of Math*, Vol. 46 (1945), p 265.

[4] A WALD, "An essentially complete class of admissible decision-functions," *Annals of Math. Stat.*, Vol 18 (1947), p 549.

[5] J. NEYMAN, "On a statistical problem arising in routine analysis and in sampling inspection of mass production," *Annals of Math Stat*, Vol. 12 (1941), p 46

[6] H. SCHEFFÉ, "On the theory of testing composite hypotheses with one constraint," *Annals of Math Stat.*, Vol. 13 (1942), p 280.

[7] E. LEHMANN, "On optimum tests of composite hypotheses with one constraint," *Annals of Math Stat*, Vol 18 (1947), p 473

[8] E LEHMANN AND H SCHEFFÉ, "On the problem of similar regions," *Proc Nat. Acad Sci*, Vol 33 (1947), p 382

[9] A. Wald, "On the power function of the analysis of variance test," *Annals of Math. Stat.*, Vol. 13 (1942), p 434.
[10] P L. Hsu, "On the power function of the E^2-test and the T^2-test," *Annals of Math. Stat*, Vol 16 (1945), p 278
[11] H K. Nandi, "On the average power of test criteria," *Sankhyā*, Vol. 8 (1946), p 67
[12] W Feller, "Note on regions similar to the sample sapce," *Stat. Res Memoirs*, Vol 2 (1938), p. 117
[13] P. L Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, Vol 32 (1941), p. 62
[14] S Kolodzieczyk, "On an important class of statistical hypotheses," *Biometrika*, Vol. 27 (1935), p. 161.
[15] P. C Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res Memoirs*, Vol. 2 (1938), p. 126.
[16] A. Wald, "Foundations of a general theory of sequential decision functions," *Econometrica*, Vol. 15 (1947) p 279.

SYMBOLIC MATRIX DERIVATIVES

By Paul S Dwyer and M S Macphail

University of Michigan and Queen's University

Summary. Let X be the matrix $[x_{mn}]$, t a scalar, and let $\partial X/\partial t$, $\partial t/\partial X$ denote the matrices $[\partial x_{mn}/\partial t]$, $[\partial t/\partial x_{mn}]$ respectively. Let $Y = [y_{pq}]$ be any matrix product involving X, X' and independent matrices, for example $Y = AXBX'C$ Consider the matrix derivatives $\partial Y/\partial x_{mn}$, $\partial y_{pq}/\partial X$. Our purpose is to devise a systematic method for calculating these derivatives. Thus if $Y = AX$, we find that $\partial Y/\partial x_{mn} = AJ_{mn}$, $\partial y_{pq}/\partial X = A'K_{pq}$, where J_{mn} is a matrix of the same dimensions as X, with all elements zero except for a unit in the m-th row and n-th column, and K_{pq} is similarly defined with respect to Y. We consider also the derivatives of sums, differences, powers, the inverse matrix and the function of a function, thus setting up a matrix analogue of elementary differential calculus This is designed for application to statistics, and gives a concise and suggestive method for treating such topics as multiple regression and canonical correlation

1. Introduction. The derivative of a matrix with respect to a scalar

$$\frac{\partial Y}{\partial x} = \frac{\partial}{\partial x}[y_{pq}] = \left[\frac{\partial y_{pq}}{\partial x}\right] \tag{1}$$

is well known and commonly used. The symbolic derivative obtained by applying a matrix of differential operators to a scalar

$$\frac{\partial y}{\partial X} = \left[\frac{\partial}{\partial x_{mn}}\right] y = \left[\frac{\partial y}{\partial x_{mn}}\right] \tag{2}$$

is not in such general use though some authors give special cases For example, if A is a symmetric matrix and X a column matrix, so that $y = X'AX$ is a quadratic form, Fraser, Duncan and Collar [1, p 48] write

$$\begin{bmatrix} \partial/\partial x_1 \\ \partial/\partial x_2 \\ \\ \partial/\partial x_n \end{bmatrix} y = 2AX \tag{3}$$

to indicate concisely the result of differentiating y with respect to the elements x_i of X.

It is to be noted that the matrix in (1) has the same dimensions (numbers of rows and columns) as the matrix Y, while the matrix in (2) has the dimensions of the matrix X.

We present an illustration of each of these types of symbolic matrix derivatives in order to clarify the concepts. Thus if

$$Y = \begin{bmatrix} x & 2x^3 & 3x^{-4} \\ e^x & \sin x & \log_e x \end{bmatrix},$$

we have

$$\frac{\partial Y}{\partial x} = \begin{bmatrix} 1 & 6x^2 & -12x^{-5} \\ e^x & \cos x & x^{-1} \end{bmatrix},$$

while if $y = x_{11}x_{32} - x_{31}x_{12}$ and

$$X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix},$$

we have

$$\frac{\partial y}{\partial X} = \begin{bmatrix} x_{32} & -x_{31} \\ 0 & 0 \\ -x_{12} & x_{11} \end{bmatrix}.$$

Suppose Y is any matrix product involving X, X' and independent matrices, for example, $Y = AXBX'C$. We may fix an element x_{mn} of X and form the matrix

$$\frac{\partial Y}{\partial x_{mn}}, \tag{4}$$

or we may fix an element y_{pq} of Y and form the matrix

$$\frac{\partial y_{pq}}{\partial X}. \tag{5}$$

The purpose of this paper is to devise a systematic method for calculating these matrices, and to give various applications in the general field of statistics.

By way of introduction we take the matrix product $Y = AX$ where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \quad \text{and } X = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{bmatrix},$$

so that

$$Y = \begin{bmatrix} a_{11}x_{11} + a_{12}x_{21} + a_{13}x_{31} & a_{11}x_{12} + a_{12}x_{22} + a_{13}x_{32} \\ a_{21}x_{11} + a_{22}x_{21} + a_{23}x_{31} & a_{21}x_{12} + a_{22}x_{22} + a_{23}x_{32} \end{bmatrix}.$$

We have then

$$\frac{\partial Y}{\partial x_{11}} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & 0 \end{bmatrix}, \qquad \frac{\partial Y}{\partial x_{12}} = \begin{bmatrix} 0 & a_{11} \\ 0 & a_{21} \end{bmatrix},$$

$$\frac{\partial Y}{\partial x_{21}} = \begin{bmatrix} a_{12} & 0 \\ a_{22} & 0 \end{bmatrix}, \qquad \frac{\partial Y}{\partial x_{22}} = \begin{bmatrix} 0 & a_{12} \\ 0 & a_{22} \end{bmatrix},$$

$$\frac{\partial Y}{\partial x_{31}} = \begin{bmatrix} a_{13} & 0 \\ a_{23} & 0 \end{bmatrix}, \qquad \frac{\partial Y}{\partial x_{32}} = \begin{bmatrix} 0 & a_{13} \\ 0 & a_{23} \end{bmatrix}.$$

These six equations can be combined in the single one

$$\text{(6)} \qquad \frac{\partial Y}{\partial x_{mn}} = AJ_{mn}$$

where J_{mn} is a matrix having dimensions of X, with all elements zero except for a unit element in the m-th row and n-th column. Similarly we find

$$\frac{\partial y_{11}}{\partial X} = \begin{bmatrix} a_{11} & 0 \\ a_{12} & 0 \\ a_{13} & 0 \end{bmatrix}, \qquad \frac{\partial y_{12}}{\partial X} = \begin{bmatrix} 0 & a_{11} \\ 0 & a_{12} \\ 0 & a_{13} \end{bmatrix},$$

$$\frac{\partial y_{21}}{\partial X} = \begin{bmatrix} a_{21} & 0 \\ a_{22} & 0 \\ a_{23} & 0 \end{bmatrix}, \qquad \frac{\partial y_{22}}{\partial X} = \begin{bmatrix} 0 & a_{21} \\ 0 & a_{22} \\ 0 & a_{23} \end{bmatrix}.$$

These four equations can be combined in the single one

$$\text{(7)} \qquad \frac{\partial y_{pq}}{\partial X} = A'K_{pq},$$

where K_{pq} is the matrix having the dimensions of Y with all elements zero except for a unit element in the p-th row and q-th column.

It should be noted that the matrices on the left of (6) and (7) are matrices composed of the basic elements $\dfrac{\partial y_{pq}}{\partial x_{mn}}$.

Other types of symbolic matrix derivatives could be defined and studied. We have selected these two main types because of their application to regression and correlation theory. The second type is more specifically indicated in the applications but the relations between the types are such that a simultaneous treatment seems appropriate.

2. Notation. Capital letters are used for matrices and small letters for scalars. It is understood that $Y, U, V, \cdots$ are matrices whose elements are functions of the elements x_{mn} of X and that $A, B, \cdots$ (unless otherwise stated)

are matrices whose elements are not functions of x_{mn}. In the development of the formulas it is understood that the differentiation is carried out with respect to x_{mn} or X. The matrix function differentiated is called Y.

We have already defined J_{mn} as the matrix having the dimensions of X with all elements zero except for a unit element in the m-th row and the n-th column, and we define K_{pq} similarly with respect to Y. We now define J'_{nm} as the matrix having the dimensions of X' with all elements zero except for a unit element in the n-th row and the m-th column, and we define K'_{qp} similarly with respect to Y'. All the formulas we obtain for $\frac{\partial Y}{\partial x_{mn}}$ involve J_{mn} or J'_{nm} while all those for $\frac{\partial y_{pq}}{\partial X}$ involve K_{pq} or K'_{qp}.

3. Differentiation of a constant. If $Y = A = [a_{pq}]$ we have at once

$$\frac{\partial y_{pq}}{\partial x_{mn}} = 0.$$

It follows that

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial}{\partial x_{mn}} [y_{pq}] = 0; \tag{8}$$

$$\frac{\partial y_{pq}}{\partial X} = \left[\frac{\partial}{\partial x_{mn}}\right] y_{pq} = 0, \tag{9}$$

where the zero matrix of (8) has the dimensions of A, while that of (9) has the dimensions of X.

4. Differentiation of a matrix with respect to itself. If $Y = X = [x_{pq}]$ we note that

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \frac{\partial x_{pq}}{\partial x_{mn}} = \begin{Bmatrix} 1 & (p = m, q = n) \\ 0 & (\text{otherwise}) \end{Bmatrix}.$$

It follows that

$$\begin{aligned} \frac{\partial Y}{\partial x_{mn}} &= \frac{\partial}{\partial x_{mn}} [y_{pq}] = J_{mn}, \\ \frac{\partial y_{pq}}{\partial X} &= \left[\frac{\partial}{\partial x_{mn}}\right] y_{pq} = K_{pq}. \end{aligned} \tag{10}$$

5. Differentiation of the transpose of a matrix with respect to the matrix. Let $Y = X'$, so that

$$y_{pq} = x_{qp}.$$

Then

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \frac{\partial x_{qp}}{\partial x_{mn}} = \begin{Bmatrix} 1 & (q = m,\ p = n); \\ 0 & (\text{otherwise}), \end{Bmatrix}$$

and we have

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial}{\partial x_{mn}} [y_{pq}] = J'_{nm}, \tag{12}$$

$$\frac{\partial y_{pq}}{\partial X} = \left[\frac{\partial}{\partial x_{mn}}\right] y_{pq} = K'_{qp}, \tag{13}$$

where J'_{nm}, K'_{qp} are defined as in section 2.

6. Differentiation of sums and differences of matrices. If

$$Y = U + V - W = [u_{pq} + v_{pq} - w_{pq}],$$

we have

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \frac{\partial u_{pq}}{\partial x_{mn}} + \frac{\partial v_{pq}}{\partial x_{mn}} - \frac{\partial w_{pq}}{\partial x_{mn}},$$

then

$$\begin{aligned}\frac{\partial Y}{\partial x_{mn}} &= \frac{\partial}{\partial x_{mn}} [y_{pq}] = \frac{\partial}{\partial x_{mn}} [u_{pq} + v_{pq} - w_{pq}] \\ &= \frac{\partial}{\partial x_{mn}} [u_{pq}] + \frac{\partial}{\partial x_{mn}} [v_{pq}] - \frac{\partial}{\partial x_{mn}} [w_{pq}] \\ &= \frac{\partial U}{\partial x_{mn}} + \frac{\partial V}{\partial x_{mn}} - \frac{\partial W}{\partial x_{mn}},\end{aligned} \tag{14}$$

and similarly

$$\frac{\partial y_{pq}}{\partial X} = \frac{\partial u_{pq}}{\partial X} + \frac{\partial v_{pq}}{\partial X} - \frac{\partial w_{pq}}{\partial X}. \tag{15}$$

7. General formulas for the differentiation of a two factor matrix product. Suppose U is a matrix with c rows and d columns and V is a matrix with d rows and e columns, then

$$Y = UV = [y_{pq}] = \sum_{s=1}^{d} u_{ps} v_{sq} \tag{16}$$

We have at once

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \sum_{s=1}^{d} \frac{\partial u_{ps}}{\partial x_{mn}} v_{sq} + \sum_{s=1}^{d} u_{ps} \frac{\partial v_{sq}}{\partial x_{mn}} \tag{17}$$

Now considering any fixed x_{mn} it is clear that the first term on the right of (17) is the same as the right hand term of (16) with $\frac{\partial u_{ps}}{\partial x_{mn}}$ in place of u_{ps}. The second term on the right of (17) is likewise the same as the right hand term of (16) with $\frac{\partial v_{sq}}{\partial x_{mn}}$ in place of v_{sq}. We may then write

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial U}{\partial x_{mn}} V + U \frac{\partial V}{\partial x_{mn}}. \tag{18}$$

Also considering a fixed y_{pq} we have

$$\frac{\partial y_{pq}}{\partial X} = \sum_{s=1}^{d} \frac{\partial u_{ps}}{\partial X} v_{sq} + \sum_{s=1}^{d} u_{ps} \frac{\partial v_{sq}}{\partial X}. \tag{19}$$

It is to be noted that this formula yields matrices of the proper dimensions (those of X) since $\frac{\partial u_{ps}}{\partial X}$ and $\frac{\partial v_{sq}}{\partial X}$ have the dimensions of X. These matrices, when multiplied by the scalar values v_{sq} and u_{ps} and summed, yield matrices of the desired dimensions.

8. Some properties of matrix products involving J's and K's. Before deriving formulas for the differentiation of products of specific factors, it seems wise to derive some formulas exhibiting certain relations involving the J's and K's. Consider the matrix A having c rows and d columns and the matrix X having d rows and e columns. Then $Y = AX$ is a matrix with c rows and e columns, J_{mn} one with d rows and e columns, J'_{nm} one with c rows and d columns, K_{pq} one with c rows and e columns and K'_{qp} one with e rows and c columns

It is easily seen by actual multiplication that

(20) *AJ_{mn} is a $c \times e$ matrix with all its elements zero except those of its n-th column which are those of the m-th column of A.* We omit further discussion of the dimensions of the matrices and assume that whenever a matrix product is written, the factors are comformable. Then we can show similarly that

(21) *$J_{mn}B$ is a matrix with all its elements zero except those of its m-th row, which are those of the n-th row of B.* Similar statements hold if J_{mn} is replaced by J'_{nm} or K_{pq} or K'_{qp} The rules are

(a) When J_{mn} (or J'_{nm} or K_{pq} or K'_{qp}) is the postmultiplier, the *first* subscript indicates the *column* of the other matrix which is placed in the column indicated by the *second* subscript.

(b) When J_{mn} (or J'_{nm} or K_{pq} or K'_{qp}) is the premultiplier, the *second* subscript indicates the *row* of the other matrix which is placed in the row indicated by the *first* subscript.

Notice also that

(22) *$A'K_{pq}$ is a matrix with all elements zero except those of its q-th column, which are those of the p-th column of A', or the p-th row of A.* A similar result holds if K_{pq} is replaced by K'_{qp} or J_{mn} or J'_{nm}.

9. Differentiation of specific two factor products. Let us start with $Y = AX$ where the various matrices involved have the dimensions indicated in the last section. Application of (18), (8), (10) gives

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial A}{\partial x_{mn}} X + A \frac{\partial X}{\partial x_{mn}} = 0 + AJ_{mn} = AJ_{mn}, \tag{23}$$

while application of (19), (11) yields

$$\frac{\partial y_{pq}}{\partial X} = \sum_{s=1}^{d} \frac{\partial a_{ps}}{\partial X} x_{sq} + \sum_{s=1}^{d} a_{ps} \frac{\partial x_{sq}}{\partial X} \tag{24}$$

$= \sum_{s=1}^{d} a_{ps} K_{sq}$

$= a_{p1}K_{1q} + a_{p2}K_{2q} + \cdots + a_{pd}K_{dq}$

$=$ a $c \times e$ matrix with all elements zero except those of its q-th column which are those of the p-th row of A

$= A'K_{pq}$ by (22).

Similar treatment of $Y = XB$ yields

$$\frac{\partial Y}{\partial x_{mn}} = \frac{\partial X}{\partial x_{mn}} B + X \frac{\partial B}{\partial x_{mn}} = J_{mn} B, \tag{25}$$

$$\frac{\partial y_{pq}}{\partial X} = \sum_{s} \frac{\partial x_{ps}}{\partial X} b_{sq} = \sum_{s} K_{ps} b_{sq} = K_{pq} B'. \tag{26}$$

If we treat $Y = AX'$ in a similar fashion, we get

$$\frac{\partial Y}{\partial x_{mn}} = AJ'_{nm}, \tag{27}$$

$$\frac{\partial y_{pq}}{\partial X} = K'_{qp} A, \tag{28}$$

while $Y = X'B$ yields

$$\frac{\partial Y}{\partial x_{mn}} = J'_{nm} B, \tag{29}$$

$$\frac{\partial y_{pq}}{\partial X} = BK'_{qp}. \tag{30}$$

It is to be noted that J always has the subscripts mn, and similarly we find always J'_{nm}, K_{pq}, K'_{qp}. We may therefore omit the subscripts on these letters. When we do so we shall also write

$$\frac{\partial Y}{\partial \langle X \rangle} \text{ for } \frac{\partial X}{\partial x_{mn}}, \qquad \frac{\partial \langle Y \rangle}{\partial X} \text{ for } \frac{\partial y_{pq}}{\partial X},$$

placing brackets $\langle\ \rangle$ around the matrix from which a fixed element is to be chosen. Thus if $Y = AX$, we write instead of (23) and (24)

$$\frac{\partial Y}{\partial \langle X \rangle} = AJ; \tag{23a}$$

$$\frac{\partial \langle Y \rangle}{\partial X} = A'K. \tag{24a}$$

The other results are summarized in lines 1–5 of Table I

Examination of (18) and (19) shows that the derivatives of products with two variable factors are obtained by adding the results obtained by holding

each factor constant while differentiating the other. With this in mind, (23)–(30) can be used to obtain the derivatives of double products involving X and X'. Thus if $Y = XX$, we get

$$\frac{\partial Y}{\partial \langle X \rangle} = JX + XJ, \qquad \frac{\partial \langle Y \rangle}{\partial X} = KX' + X'K. \tag{31}$$

Other double product formulas involving X and X' are given in Table I.

TABLE I

Formula	Y	$\frac{\partial Y}{\partial \langle X \rangle}$	$\frac{\partial \langle Y \rangle}{\partial X}$
1	AB	0	0
2	AX	AJ	$A'K$
3	XB	JB	KB'
4	AX'	AJ'	$K'A$
5	$X'B$	$J'B$	BK'
6	XX	$JX + XJ$	$KX' + X'K$
7	$X'X$	$J'X + X'J$	$XK' + XK$
8	XX'	$JX' + XJ'$	$KX + K'X$
9	$X'X'$	$J'X' + X'J'$	$X'K' + K'X'$

The formulas for $\frac{\partial Y}{\partial \langle X \rangle}$ are written down very easily, but those for $\frac{\partial \langle Y \rangle}{\partial X}$ are not so easy to write. However the values of $\frac{\partial Y}{\partial \langle X \rangle}$ and $\frac{\partial \langle Y \rangle}{\partial X}$ in formulas 2–5 of Table I are such that the results for $\frac{\partial \langle Y \rangle}{\partial X}$ may be obtained from those for $\frac{\partial Y}{\partial \langle X \rangle}$ with the use of a few simple rules. They are

(a) Each J becomes K and each J' becomes K'.

(b) The pre (or post) multiplier of J becomes its transpose.

(c) The pre (or post) multiplier of J' becomes a post (or pre) multiplier of K'.

These rules are immediately applicable to the double products. Thus when $Y = X'X$ we have

$$\frac{\partial Y}{\partial \langle X \rangle} = J'X + X'J,$$

and so

$$\frac{\partial \langle Y \rangle}{\partial X} = XK' + XK.$$

10. Differentiation of three (or more) factor products. Products with three factors can be differentiated by the formulas of the last section if two adjacent factors are constant. Thus if $Y = ABX$, we have

$$\frac{\partial Y}{\partial \langle X \rangle} = ABJ, \qquad \frac{\partial \langle Y \rangle}{\partial X} = B'A'K.$$

It is not yet demonstrated that these rules are applicable to the products AXB and $AX'B$ However it can be shown by the general methods indicated earlier that if $Y = AXB$, we obtain

$$\frac{\partial Y}{\partial \langle X \rangle} = AJB, \qquad \frac{\partial \langle Y \rangle}{\partial X} = A'KB', \tag{33}$$

while if $Y = AX'B$ we have

$$\frac{\partial Y}{\partial \langle X \rangle} = AJ'B, \qquad \frac{\partial \langle Y \rangle}{\partial X} = BK'A \tag{34}$$

It is now apparent that the rules of the last section apply to situations in which there are both pre and post multipliers.

The general theory for two-factor products is immediately extendable. Thus if $Y = UVW$ with $y_{pq} = \sum_s \sum_r u_{ps}v_{sr}w_{rq}$ then the basic element is

$$\frac{\partial y_{pq}}{\partial x_{mn}} = \sum_s \sum_r \frac{\partial u_{ps}}{\partial x_{mn}} v_{sr} w_{rq} + \sum_s \sum_r u_{ps} \frac{\partial v_{sr}}{\partial x_{mn}} w_{rq} + \sum_s \sum_r u_{ps} v_{sr} \frac{\partial w_{rq}}{\partial x_{mn}}, \tag{35}$$

and the formulas result from treating each factor in turn as the only variable. For example if $Y = XX'X$, we have

$$\frac{\partial Y}{\partial \langle X \rangle} = JX'X + XJ'X + XX'J, \tag{36}$$

and

$$\begin{aligned} \frac{\partial \langle Y \rangle}{\partial X} &= K(X'X)' + XK'X + (XX')'K \\ &= KX'X + XK'X + XX'K. \end{aligned} \tag{37}$$

The symbolic derivatives of certain triple product matrices are presented in Table II.

The rules are sufficiently general to take care of matrices with more than three factors. Thus if $Y = A'X'XB$, we have

$$\frac{\partial Y}{\partial \langle X \rangle} = A'J'XB + A'X'JB \tag{38}$$

and

$$\frac{\partial \langle Y \rangle}{\partial X} = XBK'A' + XAKB', \tag{39}$$

and in the special case $B = A$, we get

$$\frac{\partial Y}{\partial \langle X \rangle} = A'(J'X + X'J)A, \tag{40}$$

$$\frac{\partial \langle Y \rangle}{\partial X} = XA(K' + K)A'. \tag{41}$$

Similarly if $Y = X'A'AX$, we get

$$\frac{\partial Y}{\partial \langle X \rangle} = J'A'AX + X'A'AJ, \tag{42}$$

and

$$\frac{\partial \langle Y \rangle}{\partial X} = A'AXK' + A'AXK. \tag{43}$$

TABLE II

Formula	Y	$\dfrac{\partial Y}{\partial \langle X \rangle}$	$\dfrac{\partial \langle Y \rangle}{\partial X}$
1	ABC	0	0
2	ABX	ABJ	$B'A'K$
3	AXC	AJC	$A'KC'$
4	XBC	JBC	$KC'B'$
5	ABX'	ABJ'	$K'AB$
6	$AX'C$	$AJ'C$	$CK'A$
7	$X'BC$	$J'BC$	BCK'
8	AXX	$AJX + AXJ$	$A'KX' + X'A'K$
9	XBX	$JBX + XBJ$	$KX'B' + B'X'K$
10	XXC	$JXC + XJC$	$KC'X' + X'KC'$
11	$AX'X'$	$AJ'X' + AX'J'$	$X'K'A + K'AX'$
12	$X'BX'$	$J'BX' + X'BJ'$	$BX'K' + K'X'B$
13	$X'X'C$	$J'X'C + X'J'C$	$X'CK' + CK'X'$
14	$AX'X$	$AJ'X + AX'J$	$XK'A + XA'K$
15	$X'BX$	$J'BX + X'BJ$	$BXK' + B'XK$
16	$X'XC$	$J'XC + X'JC$	$XCK' + XKC'$
17	AXX'	$AJX' + AXJ'$	$A'KX + K'AX$
18	XBX'	$JBX' + XBJ'$	$KXB' + K'XB$
19	$XX'C$	$JX'C + XJ'C$	$KC'X + CK'X$
20	XXX	$JXX + XJX + XXJ$	$KX'X' + X'KX' + X'X'K$
21	XXX'	$JXX' + XJX' + XXJ'$	$KXX' + X'KX + K'XX$
22	$XX'X$	$JX'X + XJ'X + XX'J$	$KX'X + XK'X + XX'K$
23	$X'XX$	$J'XX + X'JX + X'XJ$	$XXK' + XKX' + X'XK$
24	$XX'X'$	$JX'X' + XJ'X' + XX'J'$	$KXX + X'K'X + K'XX'$
25	$X'XX'$	$J'XX' + X'JX' + X'XJ'$	$XX'K' + XKX + K'X'X$
26	$X'X'X$	$J'X'X + X'J'X + X'X'J$	$X'XK' + XK'X' + XXK$
27	$X'X'X'$	$J'X'X' + X'J'X' + X'X'J'$	$X'X'K' + X'K'X' + K'X'X'$

Finally if $Y = XAX'AX$, we get

$$\frac{\partial Y}{\partial \langle X \rangle} = JAX'AX + XAJ'AX + XAX'AJ, \tag{44}$$

$$\frac{\partial Y}{\partial \langle X \rangle} = KX'A'XA' + AXK'XA + A'XA'X'K. \tag{45}$$

11. Vector results. It should be emphasized that each of the above results is a general result. More specific results may be obtained in case one (or more)

of the matrices is a vector. For example if X_c is a column matrix and $Y = X_c'BX_c$, then Y is a scalar, so K and K' are both unity and we have from Table II (15)

$$\frac{\partial\langle Y\rangle}{\partial X} = BX_c + B'X_c = (B + B')X_c. \tag{46}$$

If in addition B is symmetric, $B' = B$ and we have

$$\frac{\partial\langle Y\rangle}{\partial X} = 2BX_c,$$

which is the result indicated in (3).

12. Differentiation of the inverse of X. It is possible to use implicit differentiation to derive formulas for $\frac{\partial X^{-1}}{\partial\langle X\rangle}$ and $\frac{\partial\langle X^{-1}\rangle}{\partial X}$. We write $I = XX^{-1}$ and get

$$\frac{\partial I}{\partial\langle X\rangle} = 0 = JX^{-1} + X\frac{\partial X^{-1}}{\partial\langle X\rangle},$$

so that

$$\frac{\partial X^{-1}}{\partial\langle X\rangle} = -X^{-1}JX^{-1}, \tag{47}$$

whence

$$\frac{\partial\langle X^{-1}\rangle}{\partial X} = -(X^{-1})'K(X^{-1})'. \tag{48}$$

The formula (47) is a generalization of a known matrix differential formula [3:3.4].

In a similar way we derive

$$\frac{\partial(X')^{-1}}{\partial\langle X\rangle} = -(X')^{-1}J'(X')^{-1}, \tag{49}$$

$$\frac{\partial\langle(X')^{-1}\rangle}{\partial X} = -(X')^{-1}K'(X')^{-1}. \tag{50}$$

13. Differentiation of a function of a function. The theory developed in the earlier sections is sufficiently general to be useful in differentiating a function of a function if the functions involve addition, subtraction, premultiplication, postmultiplication, and inverse. For example if

$$Y = Z'Z \quad \text{with} \quad Z = AX \tag{51}$$

we have

$$\frac{\partial Y}{\langle\partial X\rangle} = \frac{\partial Z'}{\partial\langle X\rangle}Z + Z'\frac{\partial Z}{\partial\langle X\rangle},$$

and since

$$\frac{\partial Z'}{\partial\langle X\rangle} = J'A' \quad \text{and} \quad \frac{\partial Z}{\partial\langle X\rangle} = AJ,$$

(52) $$\frac{\partial\langle Y\rangle}{\partial X} = J'A'Z + Z'AJ,$$

and thence

(53) $$\frac{\partial\langle Y\rangle}{\partial X} = A'ZK' + A'ZK.$$

These results are equivalent to those of (42) and (43).

14. Differentiation of a power of a square matrix. The values of the symbolic derivatives of X^2, X^3 with respect to X are given in Tables I and II. It can be shown similarly that if n is a positive integer

(54) $$\frac{\partial X^n}{\partial\langle X\rangle} = JX^{n-1} + \sum_{s=1}^{n-2} X^s J X^{n-s-1} + X^{n-1}J,$$

and this can be written as

(55) $$\frac{\partial X^n}{\partial\langle X\rangle} = \sum_{s=0}^{n-1} X^s J X^{n-s-1},$$

if we adopt the convention that X^0 is I. It follows at once that

(56) $$\frac{\partial\langle X^n\rangle}{\partial X} = \sum_{s=0}^{n-1} X'^s K (X')^{n-s-1}$$

It is thence possible to derive formulas for the symbolic derivatives of X^{-n}. Since $X^{-n}X^n = \text{I}$, we have

(57) $$\frac{\partial X^{-n}}{\partial\langle X\rangle} X^n + X^{-n}\left[\sum_{s=0}^{n-1} X^s J X^{n-s-1}\right] = 0,$$

so

(58) $$\frac{\partial X^{-n}}{\partial\langle X\rangle} = -X^{-n}\left[\sum_{s=0}^{n-1} X^s J X^{n-s-1}\right] X^{-n},$$

and

(59) $$\frac{\partial\langle X^{-n}\rangle}{\partial X} = -X^{-n}\left[\sum_{s=0}^{n-1} (X')^s K (X')^{n-s-1}\right] X^{-n}.$$

15. Applications. We consider the classical theory of least squares, a matrix presentation of which is available in [2]. Suppose that y and x_i are measured from their means and that y is to be estimated from the n variables x_i. Form the values of y into a column matrix Y and the values of x_i into an N by n matrix X. Introduce the column matrix B of n parameters b_i and define

(60) $$E = Y - XB.$$

Note that the matrix $E'E$ is in this case the single element matrix which is the sum of the squares of the residuals Following the least squares method we minimize this by differentiating with respect to the elements of B. We first note that

$$E'E = (Y' - B'X')(Y - XB) \tag{61}$$
$$= Y'Y - Y'XB - B'X'Y + B'X'XB$$

Then we write down first

$$\frac{\partial(E'E)}{\partial\langle B\rangle} = -Y'XJ - J'X'Y + J'X'XB + B'X'XJ, \tag{62}$$

from which we get

$$\frac{\partial\langle E'E\rangle}{\partial B} = -X'YK - X'YK' + X'XBK' + X'XBK \tag{63}$$
$$= -X'(Y - XB)(K + K') = -X'E(K + K')$$

The J's and K's are associated with B and $E'E$ respectively. Here $E'E$ is scalar so that $K = K' = 1$ and we have

$$\frac{\partial\langle E'E\rangle}{\partial B} = -2X'E. \tag{64}$$

The equation $X'E = 0$, obtained by equating the right hand side of (64) to zero, is a statement of the normal equations in matrix form.

Equation (64) may also be obtained with the use of the methods of section 13. In this case

$$\frac{\partial E}{\partial\langle B\rangle} = -XJ, \qquad \frac{\partial E'}{\partial\langle B\rangle} = -J'X',$$

and we have

$$\frac{\partial(E'E)}{\partial\langle B\rangle} = \frac{\partial E'}{\partial\langle B\rangle} E + E' \frac{\partial E}{\partial\langle B\rangle} = -J'X'E - E'XJ; \tag{65}$$

so

$$\frac{\partial\langle E'E\rangle}{\partial B} = -X'EK' - X'EK = -X'E(K' + K). \tag{66}$$

The equation (64) is also applicable to the more general problem in which y_1 and y_2 are estimated from the same set of variables x_i. The only change needed is to regard Y, B, E as two-column matrices so that $E'E$ is a matrix with two rows and columns which we denote by

$$\begin{bmatrix} \epsilon_{11} & \epsilon_{12} \\ \epsilon_{21} & \epsilon_{22} \end{bmatrix}.$$

We require $\frac{\partial \epsilon_{11}}{\partial B} = 0$ and $\frac{\partial \epsilon_{22}}{\partial B} = 0$. From equation (63), inserting subscripts, we get

$$\frac{\partial \epsilon_{11}}{\partial B} = -X'E(K_{11} + K'_{11})$$
$$= -2X'EK_{11};$$
$$\frac{\partial \epsilon_{22}}{\partial B} = -2X'EK_{22}.$$

It is easily seen that $\frac{\partial \epsilon_{11}}{\partial B} = \frac{\partial \epsilon_{22}}{\partial B} = 0$ is equivalent to $X'E = 0$, the same equation as we obtained in the last paragraph. We also arrive at the incidental result that in minimizing $\Sigma\epsilon_1^2$, and $\Sigma\epsilon_2^2$ separately we find at the same time a stationary value of $\Sigma\epsilon_1\epsilon_2$.

In this way we can treat two or more simultaneous regression problems with this general notation as easily as we can treat one.

As a second application of the theory we outline the initial steps in the direction of the formulas for canonical correlation [4], [5]. In this case A and B are unknown column vectors with X and Y known rectangular matrices. Then XA is a column matrix:

$$XA = L = \begin{bmatrix} l_1 \\ l_2 \\ \cdot \\ \cdot \\ l_N \end{bmatrix},$$

whose elements l_i may be regarded as observed values of a linear form l. Similarly $YB = \Lambda$, a column matrix whose elements may be replaced as observed values of a linear form λ. It is desired to find A and B such that l and λ may have the largest correlation coefficient, and to find the size of this coefficient. Then $A'X'XA$, $B'Y'YB$, and $B'Y'XA = A'X'YB$ are scalars, and

$$\rho = \frac{B'Y'XA}{\sqrt{(A'X'XA)(B'Y'YB)}}. \tag{67}$$

If the scales of X and Y are chosen so that $A'X'XA = 1$ and $B'Y'YB = 1$, we have

$$\rho = B'Y'XA = A'X'YB. \tag{68}$$

Using Lagrange multipliers we set

$$\phi = B'Y'XA + \frac{c}{2}(1 - A'X'XA) + \frac{d}{2}(1 - B'Y'YB), \tag{69}$$

and differentiate with respect to the elements of A and B. We first differentiate ϕ with respect to A after replacing $B'Y'XA$ by $A'X'YB$:

(70) $$\frac{\partial\phi}{\partial\langle A\rangle} = J'X'YB - \frac{c}{2}(J'X'XA + A'X'XJ);$$

(71) $$\frac{\partial\langle\phi\rangle}{\partial A} = X'YBK' - \frac{c}{2}(X'XAK' + X'XAK)$$

(The J's and K's are associated with A and ϕ respectively) We set $\frac{\partial\langle\phi\rangle}{\partial A} = 0$ with $K = K' = 1$ to get

(72) $$X'YB = cX'XA,$$

whence by (57)

(73) $$\rho = A'X'YB = cA'X'XA = c,$$

and

(74) $$X'YB = \rho X'XA.$$

Similar differentiation with respect to B gives $\rho = d$ and

(75) $$Y'XA = \rho Y'YB.$$

The further steps in the development of canonical correlation theory are based on (74) and (75).

A third application is to orthogonal regression. The situation is very similar to that of the first illustration, but the errors are measured orthogonal to the plane of best fit. As before we take the variates as measured from their means and so have the basic equation

(76) $$D = \frac{b_1x_1 + b_2x_2 + \cdot\cdot + b_kx_k}{\sqrt{b_1^2 + b_2^2 + \cdots + b_k^2}}.$$

This can be written as

(77) $$D = l_1x_1 + l_2x_2 + \cdots + l_kx_k = XL \text{ with } L'L = 1.$$

It follows that the quantity to be minimized is

(78) $$D'D = L'X'XL.$$

With the use of Lagrange multipliers we have

(79) $$\Phi = L'X'XL + \lambda(1 - L'L)$$

so that

(80) $$\frac{\partial\phi}{\partial\langle L\rangle} = J'X'XL + L'X'XJ - \lambda(J'L + L'J),$$

(81) $$\frac{\partial\langle\phi\rangle}{\partial L} = X'XLK' + X'XLK - \lambda(LK' + LK)$$

from which

(82) $$2X'XL - 2\lambda L = 0$$

and the values can be determined from the equation

$$(X'X - \lambda)L = 0. \tag{83}$$

The solution continues with the use of the characteristic equation.

It is to be noted from (79) and (82) that

$$D'D = L'X'XL = \lambda L'L = \lambda$$

so that (83) becomes

$$(X'X - D'D)L = 0. \tag{84}$$

A fourth illustration uses symbolic derivatives in obtaining the principal components of a total variance [5,252] The variable portion of the exponent of the multivariate normal can be written $Y'AY$ where Y is the column vector $[y_1, \cdots, y_k]$ and A is a k by k matrix. We set this equal to a constant, say C, and get the equation of the k dimensional ellipsoid It is desired to locate the extrema of this ellipsoid. To do this we find the extrema of $Y'Y$. Using the Lagrange multiplier we have

$$\phi = Y'Y + \lambda(C - Y'AY) \tag{85}$$

so that

$$\frac{\partial\phi}{\partial\langle Y\rangle} = J'Y + Y'J - \lambda(J'AY + Y'AJ), \tag{86}$$

$$\frac{\partial\langle\phi\rangle}{\partial Y} = YK' + YK - \lambda(AYK' + AYK), \tag{87}$$

so that there results

$$Y - \lambda AY = 0. \tag{88}$$

Pre-multiplying by A^{-1} we get

$$(A^{-1} - \lambda)Y = 0 \tag{89}$$

and pre-multiplying by Y' gives the important relation

$$Y'Y = \lambda C. \tag{90}$$

A fifth illustration utilizes symbolic differentiation in developing the theory of the linear discriminant function [6, 341] [8, 124]. As in the other illustrations, the variates are measured about their means. The unknown multipliers are indicated by the vector L. Then

$$Z = XL \tag{91}$$

is the general matrix equation while

$$\begin{aligned} Z_1 &= X_1L \\ Z_2 &= X_2L \end{aligned} \tag{92}$$

are the corresponding equations for the two groups. Then

$$\bar{Z}_1 = \bar{X}_1 L,\ \bar{Z}_2 = \bar{X}_2 L, \text{ and } \bar{Z}_1 - \bar{Z}_2 = (\bar{X}_1 - \bar{X}_2)L = DL, \tag{93}$$

$$\begin{aligned} Z_1 - \bar{Z}_1 &= (X_1 - \bar{X}_1)L = Y_1 L, \\ Z_2 - \bar{Z}_2 &= (X_2 - \bar{X}_2)L = Y_2 L \end{aligned} \tag{94}$$

The within group variation, $L'Y_1'Y_1L + L'Y_2'Y_2L$, is then divided into the between group variation, $L'D'DL$, to get

$$G = \frac{L'D'DL}{L'Y_1'Y_1'L + L'Y_2'Y_2L} = \frac{A}{B}. \tag{95}$$

We wish to maximize G. Since A and B are scalars $\frac{\partial\langle G\rangle}{\partial L} = 0$ reduces to

$$\frac{\partial\langle B\rangle}{\partial L} = \frac{1}{G}\frac{\partial\langle A\rangle}{\partial L} \tag{96}$$

which becomes, with further differentiation

$$(Y_1'Y_1 + Y_2'Y_2)L = D'\left(\frac{DL}{G}\right). \tag{97}$$

Since $\frac{DL}{G}$ is a scalar, we have

$$(Y_1'Y_1 + Y_2'Y_2)L = cD. \tag{98}$$

Any convenient value of c can be used for purposes of discrimination. It is customary to take $c = 1$ and then to adjust (98) so that some l_i is unity.

A final illustration applies symbolic matrix differentiation to a theorem of multiple factor analysis. This presentation parallels that given by Thurstone [7,473–477] for transforming any factorial matrix into a principal axes matrix. The matrix

$$F = [a_{ij}] \tag{99}$$

has p rows and r columns, $r \leq p$, such that

$$FF' = R \tag{100}$$

where R is a $p \times p$ correlation matrix.

It is desired to apply the unitary orthogonal transformation L to F in such a way as to produce a matrix, called F_p, which has the sums of the squares in respective columns a maximum. This can be done by maximizing simultaneously the diagonal terms of $F_p'F_p$ where

$$F_p = FL. \tag{101}$$

Again using Lagrange multipliers, we have

$$\phi = L'F'FL + \lambda(I - L'L). \tag{102}$$

This equation has the same analytical form as (79). Differentiation leads to the result

$$(F'F - \lambda)L = 0. \tag{103}$$

The solution of (103) gives the value L which can be substituted in (101) to obtain F_p.

14. Conclusion. Two types of symbolic matrix derivatives have been defined. Laws have been developed for the basic operations of addition, subtraction, multiplication, inverse, and powers. Laws for more extended functions can be worked out on the basis of principles enunciated.

Applications are given to certain multivariate problems. It is our thesis that with these differentiation formulas available, much work in multivariate analysis can be carried on with a simple matrix notation.

REFERENCES

[1] R. A. FRAZER, W. J. DUNCAN, AND A. R. COLLAR, *Elementary Matrices and Some Applications to Dynamics and Differential Equations*, Cambridge University Press, 1936.

[2] P. S. DWYER, "A matrix presentation of least squares and correlation theory," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 82–89.

[3] A. D. MICHAL, *Matrix and Tensor Calculus*, New York, John Wiley and Sons, 1947.

[4] H. HOTELLING, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 321–377.

[5] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1940.

[6] M. G. KENDALL, *Advanced Theory of Statistics*, Vol. II, London, Charles Griffin and Co. Ltd., 1946.

[7] L. L. THURSTONE, *Multiple Factor Analysis*, Chicago, University of Chicago Press, 1947.

[8] P. G. HOEL, *Introduction to Mathematical Statistics*, New York, John Wiley and Sons, 1947.

ON THE LIMITING DISTRIBUTIONS OF ESTIMATES BASED ON SAMPLES FROM FINITE UNIVERSES[1]

By William G. Madow

Institute of Statistics, University of North Carolina

1. Summary. The paper shows that under very broad conditions the usual theorems concerning the limiting distributions of estimates hold for estimates based on samples selected from finite universes, at random without replacement. It may be remarked that under the same conditions, the same conclusions are true for random sampling from finite universes with replacement, if the universes are permitted to change within the limitations set by condition W.

2. Introduction. It has long been known that the limiting distribution of arithmetic means of samples selected at random with replacement from finite universes, or from infinite universes is normal under very general conditions When, however, a sample is selected from a finite universe without replacement, and the size of the sample as compared with that of the universe is too large for the universe to be treated as infinite, the proof that the limiting distribution of the mean is normal appears to have been given only for the case where the universe is multinomial[2] In this paper we prove that the limiting distribution of the mean is normal provided only that as the universe increases in size, the higher moments do not increase too rapidly as compared with the variance, and that for sufficiently large sizes of sample and population the ratio of size of sample to size of universe is bounded away from 1. Various extensions are given, but these are almost immediate consequences of the theorem on the limiting distribution of the mean.

The method used is that of showing that the moments of the standardized mean tend to those of the normal distribution. In doing this we generalize a theorem of Wald and Wolfowitz,[3] by making it applicable to permutations of samples from finite populations, and by reducing a little the conditions on the coefficients The theorem on the mean is then a simple corollary.

We also note that with these proofs on limiting distributions we can make the corresponding assertions concerning characteristic functions. Although no applications of this fact are given, it seems likely that some useful results could be obtained.

3. Preliminary lemmas. In calculating the k-th moments and their limits we

[1] Presented to the American Mathematical Society at a meeting held in New York City on April 17, 1948.

[2] See F. N. David, "Limiting distributions connected with certain methods of sampling human populations," *Stat Res. Mem*, Vol. 2 (1938), pp. 69–90, especially p. 77.

[3] A Wald and J. Wolfowitz, "Statistical tests based on permutations of the observations," *Annals of Math. Stat.*, Vol. 5 (1944), pp 358–372, especially p 359.

shall use an infrequently given form of the multinomial expansion and some properties of symmetric polynomials. In this section we make the necessary definitions, and present four lemmas embodying the results we shall use.[4]

A t-partition of a positive integer k consists of t positive integers $\alpha_1, \cdots, \alpha_t$ such that $\alpha_1 + \cdots + \alpha_t = k$. Two t-partitions $\alpha_1, \cdots, \alpha_t$ and $\beta_1, \cdots, \beta_t$ of k will be said to be distinct if for at least one value of h we have $\alpha_h \neq \beta_h$

Let $\varphi(\alpha_1, \cdots, \alpha_t)$, written $\varphi(\alpha)$, be any function of the t-partitions of k. By $\Sigma_{1t}\varphi(\alpha)$ we shall mean the summation of $\varphi(\alpha_1, \cdots, \alpha_t)$ over all distinct t-partitions of k.

By $\Sigma_{2t}\varphi(\alpha)$ we shall mean the summation of $\varphi(\alpha)$ over all distinct permutations of $\alpha_1, \cdots, \alpha_t$.

By $\Sigma_{3t}\varphi(\alpha)$ we shall mean the summation of $\varphi(\alpha)$ over all distinct t partitions of k satisfying the condition $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_t$.

Let $\psi(\nu_1, \cdots, \nu_t)$ be any function of the variables $\nu_1, \cdots, \nu_t$. Then by $\Sigma_{4n}\psi(\nu_1, \cdots, \nu_t)$ we shall mean the summation of $\psi(\nu_1, \cdots, \nu_t)$ over all possible selections of t integers from 1 to n arranged so that $\nu_1 > \nu_2 > \cdots > \nu_t$.

The formula for the multinomial given below is not presented as a new result It is given only as a means of referring to the result we need.

LEMMA 1. *Let $\xi_1, \cdots, \xi_n$ be any quantities or random variables and let k be a positive integer. Then*

$$(\xi_1 + \cdots + \xi_n)^k = \sum_{t=1}^{k} \Sigma_{1t}\, C^k_{\alpha_1 \cdots \alpha_t} \Sigma_{4n}\, \xi_{\nu_1}^{\alpha_1} \cdots \xi_{\nu_t}^{\alpha_t},$$

where

$$C^k_{\alpha_1 \cdots \alpha_t} = \frac{k!}{\alpha_1! \cdots \alpha_t!}.$$

The proof is omitted.

The following lemma will be useful in connection with several of the results of this section:

LEMMA 2. *If $\varphi(\alpha)$ is a function of the t-partitions of k, then*

$$\Sigma_{1t}\varphi(\alpha) = \Sigma_{3t}\Sigma_{2t}\varphi(\alpha).$$

The verbalization of the lemma is practically its proof.

Let us now define certain symmetric polynomials that we shall use

Let $S_{\alpha_1, \cdots, \alpha_t} = \Sigma \xi_{\nu_1}^{\alpha_1} \cdots \xi_{\nu_t}^{\alpha_t}$ where the α's are positive integers and the summation extends over all possible arrangements $\nu_1, \cdots, \nu_t$ of t of the integers $1, \cdots, N$. Hence there will be $N^{(t)} = N(N-1) \cdots (N-t+1)$ terms in $S_{\alpha_1, \cdots, \alpha_t}$.

LEMMA 3 *Suppose that $t_1, \cdots, t_h$ are an h partition of t, that*

$$\alpha_{t_1 + \cdots + t_{i-1}+1} = \cdots = \alpha_{t_1 + \cdots + t_i}, \quad (i = 1, \cdots, h;\ t_0 = 0),$$

[4] The order of sections 3 and 4 is largely a matter of taste; some may prefer to treat section 3 as an appendix to section 4 to be referred to when necessary.

and that

$$\alpha_1 \neq \alpha_{t_1+1} \neq \cdots \neq \alpha_{t_1+\cdots+t_{h-1}+1}.$$

Then, defining

$$S'_{\alpha_1 \cdots \alpha_t} = \Sigma_{2t}\Sigma_{1n}\xi_{\nu_1}^{\alpha_1} \cdots \xi_{\nu_t}^{\alpha_t}, \tag{3.1}$$

it follows that

$$S_{\alpha_1, \cdots, \alpha_t} = t_1! \cdots t_h! S'_{\alpha_1, \cdots, \alpha_t}.$$

To prove Lemma 3, it is only necessary to note that each term of $S'_{\alpha_1, \cdots, \alpha_t}$ will determine $t_1! \cdots t_h!$ equal terms of $S_{\alpha_1, \cdots, \alpha_t}$.

Although the moments that we shall obtain will be functions of $S_{\alpha_1, \cdots, \alpha_t}$, the condition that we shall use on the moments can be interpreted directly only in terms of S_j. Consequently, in order to be able to analyze the implications of that condition on $S_{\alpha_1, \cdots, \alpha_t}$, we state the following lemma:

LEMMA 4. *The symmetric polynomial $S_{\alpha_1, \cdots, \alpha_t}$ is equal to a sum of products of the form*

$$\pm S_{\gamma_1} S_{\gamma_2} \cdots S_{\gamma_h}$$

where $\gamma_1, \cdots, \gamma_h$ are an h-partition of k, $h \leq t$, and each γ is a sum of one or more of the α's. Furthermore, if $S_1 = 0$, then $h \leq [k/2]$ where $[k/2] = k/2$ if k is even and $[k/2] = (k-1)/2$ if k is odd. This follows from the result

$$S_{\alpha_t} \cdot S_{\alpha_1, \cdots, \alpha_{t-1}} = S_{\alpha_1, \cdots, \alpha_t} + S_{\alpha_1+\alpha_t, \alpha_2, \cdots, \alpha_{t-1}} + \cdots + S_{\alpha_1, \cdots, \alpha_{t-2}, \alpha_{t-1}+\alpha_t}. \tag{3.2}$$

PROOF: It is easy to prove (3.2) by comparing terms. Then the other assertions follow from the repeated use of (3.2) and the resulting fact that each γ is a sum of one or more of the α's.

4. The limiting distribution.

In this section we obtain the generalization of the theorem of Wald and Wolfowitz to which reference was made above.

Let $U_1, U_2, \cdots, U_N, \cdots$ be a sequence of universes, the universe U_N containing the elements[5] $x_{\nu N}$ and let the arithmetic mean of the elements of U_N be denoted by $\bar{x}_N$. Furthermore, let

$$\mu_{rN} = \mu_r(U_N) = \left(\frac{1}{N}\right) \sum_{\nu} (x_{\nu N} - \bar{x}_N)^r.$$

Let $C_1, C_2, \cdots, C_n, \cdots$ be a sequence of sets of coefficients, the set C_n containing the elements c_{jn} and let the arithmetic mean of the elements of C_n be denoted by $\bar{c}_n$. We exclude the possibility that the elements of any C_n all vanish, and hence we can suppose that $\sum_j c_{nj}^2 = 1$. Furthermore, let

[5] The letter ν will assume all integral values from 1 to N. The letter r will assume all positive integral values. The letter j will assume all integral values from 1 to n. The letter t will assume all integral values from 1 to k. The symbol lim will stand for the limit as n or N or both, as the case may be, increase without limit, it being understood that $\lim n/N < 1$.

$$\mu'_{rn} = \mu'_r(C'_n) = \left(\frac{1}{n}\right) \sum_j c^r_{jn} .$$

Since $\sum_j (c_{jn} - \bar{c}_n)^2 > 0$, it follows that, if we define $A_n = n^{1/2}\bar{c}_n$, then $A_n^2 \leq 1$.

Let n elements be selected at random without replacement from U_N and let us denote these elements by x'_{jn}, the subscript j indicating the order of selection, i.e., x'_{in} is the i-th element of U_N selected for the sample even though it may be x_{NN}.

The linear function that we shall study is

$$z_n = c_{1n}x'_{1N} + \cdots + c_{nn}x'_{nN} ,$$

i.e., the value of z_n is determined by multiplying the j-th element selected for the sample by c_{jn} and summing for j. Then, since $Ex'_{in} = \bar{x}_N$, we have

$$Ez_n = n\bar{x}_N\bar{c}_n .$$

Furthermore,

$$\sigma^2_{z_n} = \left(\frac{N}{N-1}\right) \mu_{2N} \left(1 - \frac{n}{N} A_n^2\right).$$

To see this we first note that

$$\sum_{i \neq j=1}^{n} c_{in}\, c_{jn} = n^2 \bar{c}_n^2 - 1,$$

$$E(x'_{in} - \bar{x}_N)^2 = \mu_{2N} ,$$

and, if $i \neq j$,

$$E(x'_{in} - \bar{x}_N)(x'_{jn} - \bar{x}_N) = -\mu_{2N} \left(\frac{1}{N-1}\right).$$

From the definition of variance we have

$$\sigma^2_{z_n} = E(z_n - Ez_n)^2 = \sum_{i,j=1}^{n} c_{in}\, c_{jn}\, E(x'_{in} - \bar{x}_N)(x'_{jn} - \bar{x}_N),$$

and making the indicated substitutions the result follows from a few simple manipulations.

If we define $\bar{x}_n$ to be the arithmetic mean of $x'_{1N}, \cdots, x'_{nN}$, then it follows that $\sqrt{n}\, c_{jn} = 1$ and, as is well known,

$$E\bar{x}_n = \bar{x}$$

$$\sigma^2_{\bar{x}_n} = \left(\frac{N-n}{N-1}\right) \frac{\mu_{2n}}{n} .$$

Hence, if we can find the limiting distribution of

$$Z_n = \frac{z_n - Ez_n}{\sigma_{z_n}} ,$$

then the limiting distribution of $(\bar{x} - \bar{x})/\sigma_{\bar{x}}$ will be a special case

We shall need to place some sort of limitation on the sequences U_N and C_n if we are to obtain theorems on limiting distributions of statistics based on them

The condition W that we shall use is satisfied by a slightly larger class of sequences U_N and C_n than that of Wald and Wolfowitz because it does not rule out the possibility that all the elements of C_n should be equal It should be noted, however, that for their purposes this extension of the class of sequences satisfying U_N and C_n is vacuous since they required $n = N$, so that in their case if all the elements of C_n were equal, say k/N, we would have $z_N = k\,\bar{x}_N$ no matter in what order the elements of U_N were selected for the sample.

CONDITION W. The sequence U_N and C_n will satisfy the condition W if

$$\mu_{rN} = \mu_{2N}^{r/2}\,\lambda_r(N),$$

$$\mu'_{rn} = n^{-r/2}\,\lambda'_r(n),$$

and

$$\frac{nA_n^2}{N} < 1 - \epsilon,$$

for sufficiently large n and N, where a finite value λ exists such that for all r

$$\sup |\lambda_r(N)| < \lambda,$$

$$\sup |\lambda'_r(n)| < \lambda,$$

and $\epsilon > 0$.

(Note that if W is satisfied for all even values of r then W is also satisfied for all odd values of r since $\mu_{r+2}\mu_r \geq \mu_{r+1}^2$).

A general theorem on moments is the following:

THEOREM 1. *Let $S_{\alpha_1,\cdots,\alpha_t}$ and $S'_{\alpha_1,\cdots,\alpha_t}$ be defined in terms of $x_N - \bar{x}_N$ instead of ξ_ν and let $T'_{\alpha_1,\cdots,\alpha_t}$ be the same function of the c_{jn} that $S'_{\alpha_1,\cdots,\alpha_t}$ is of the $x_{\nu N} - \bar{x}_N$. Furthermore, let $E_k = EZ_n^k$. Then*

$$E_k = \sum_t \sum\nolimits_{3t} C^k_{\alpha_1\cdots\alpha_t}\,\frac{S_{\alpha_1\cdots\alpha_t}\,T'_{\alpha_1\cdots\alpha_t}}{N^{(t)}\,\sigma^k_{z_n}}. \tag{4.1}$$

PROOF: From the definition of Z_n and Lemma 1, it follows that

$$\sigma^k_{z_n} E_k = \sum_t \sum\nolimits_{1t} C^k_{\alpha_1\cdots\alpha_t} \sum\nolimits_{4n} c^{\alpha_1}_{\nu_1 n}\cdots c^{\alpha_t}_{\nu_t n}\, E(x'_{\nu_1 N} - \bar{x}_N)^{\alpha_1}\cdots(x'_{\nu_t N} - \bar{x}_N)^{\alpha_t}.$$

Since we are selecting at random without replacement it follows that

$$N^{(t)} E(x'_{\nu_1 N} - \bar{x}_N)^{\alpha_1}\cdots(x'_{\nu_t N} - \bar{x}_N)^{\alpha_t} = S_{\alpha_1\cdots\alpha_t}.$$

If we now use Lemma 2 to replace Σ_{1t} by $\Sigma_{3t}\Sigma_{2t}$, we then obtain

$$\sigma^k_{z_n} N^{(t)} E_k = \sum_t \sum\nolimits_{3t} C^k_{\alpha_1\cdots\alpha_t} S_{\alpha_1\cdots\alpha_t} \sum\nolimits_{2t} \sum\nolimits_{4n} c^{\alpha_1}_{\nu_1 n}\cdots c^{\alpha_t}_{\nu_t n},$$

since both $C^k_{\alpha_1,\cdots,\alpha_t}$ and $S_{\alpha_1,\cdots,\alpha_t}$ are invariant under permutations of $\alpha_1, \cdots, \alpha_t$. Then from (3 1) and the definition of $T'_{\alpha_1,\cdots,\alpha_t}$, it follows that (4 1) is proved.

Our fundamental theorem is:

THEOREM 2. *If the sequences U_N and C_n satisfy the condition W, then*

$$\lim E_{2j+1} = 0,$$

and

$$\lim E_{2j} = \frac{(2j)!}{2^j \cdot j!},$$

so that, for any a,

$$\lim P\{Z_n < a\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{(-x^2/2)}\, dx,$$

PROOF: We wish to show that lim E_k exists and has the values given above First consider the parts of the typical term of E_k that depend on n and N, i e., the expression

$$B = \frac{S_{\alpha_1, \cdots, \alpha_t} T'_{\alpha_1, \cdots, \alpha_t}}{N^{(t)} \mu_{2N}^{k/2} (N/N - 1)^{k/2} (1 - nA_n^2/N)^{k/2}}.$$

Since lim E_k will be the sum of the limits of a finite number of these terms, let us first determine under what conditions B will tend to zero as n and N become infinite.

From Lemma 4 it follows that

$$S_{\alpha_1, \cdots, \alpha_t} = \Sigma \pm S_{\gamma_1} S_{\gamma_2} \cdots S_{\gamma_h},$$

where $\gamma_1 + \cdots + \gamma_h = \alpha_1 + \cdots + \alpha_t$ and each of the γ's is the sum of one or more of the α's. From the definition of $S_{\alpha_1, \cdots, \alpha_t}$ in terms of $x_{\nu N} - \bar{x}_N$ it follows that $S_1 = 0$. Hence the minimum value of all γ's in any non-vanishing term of the summation is 2. Consequently we can say that for all non-vanishing terms $h \leq [k/2]$ and $h \leq t$. Finally if condition W is satisfied then

$$S_{\gamma_1} \cdots S_{\gamma_h} = N^h \mu_{2N}^{k/2} \lambda_h(N)$$

where

$$\sup |\lambda_h(N)| < \lambda^h.$$

Similarly

$$T'_{\alpha_1, \cdots, \alpha_t} = \Sigma \pm T_{\gamma_1}, \cdots, T_{\gamma_g},$$

where it may be that $T_1 \neq 0$ so that we cannot require $g \leq [k/2]$ for the term $T_{\gamma_1} \cdots T_{\gamma_g}$ to be non-vanishing. We still have, however, from Lemma 4 that $g \leq t$.

If condition W is satisfied, then

$$T_{\gamma_1} \cdots T_{\gamma_g} = n^{g-k/2} \lambda'_g(n),$$

where

$$\sup |\lambda_g'(n)| < \lambda^g.$$

Hence, from Lemma 4, the definitions of μ_{jN} and μ_{jn}' and condition W it follows that B is a sum (the number of terms does not depend on n or N) of terms like

$$D = \frac{N^h n^{g-k/2} \bar{\lambda}(N)\bar{\lambda}'(n)}{N^{(t)}(N/N - 1)^{k/2}(1 - nA_n^2/N)^{k/2}},$$

where

$$h \leq [k/2], \qquad h \leq t, \qquad g \leq t,$$

and

$$\sup |\bar{\lambda}(N)| < \infty,$$

$$\text{and} \sup |\bar{\lambda}'(n)| < \infty.$$

Since $h \leq t$, it follows that if $g < k/2$ then $\lim D = 0$ Hence, a possibly non-vanishing term must have $g \geq k/2$ and hence $t \geq k/2$ because $t \geq g$. Furthermore, $t \geq g + h - k/2$, since $h - k/2 \leq 0$ and $t \geqq g$ Hence $t - h \geq g - k/2$ Now, we can write

$$D = \frac{n^{g-k/2}}{N^{t-h}} \bar{\lambda}(N, n),$$

where

$$\sup |\bar{\lambda}(N, n)| < \infty,$$

since $nA_n^2/N < 1 - \epsilon$ for sufficiently large n and N.
Hence

$$\lim D = 0,$$

unless, perhaps, when $g - k/2 = t - h$, i e., $h - k/2 = t - g$. Since $h - k/2 \leq 0$ and $t - g \geq 0$, *it follows that we must have* $h = k/2$ *and* $t = g$ *for* $\lim D$ *to be possibly not zero.*

If k is odd, then $h \leq (k - 1)/2$ and hence

$$\lim E_{2j+1} = 0,$$

since all terms obtained by expanding it as above will tend to zero

If k is even, say $k = 2j$, and $\lim D$ is possibly non-vanishing, then h must equal j and we must have $\gamma_1 = \cdots = \gamma_j = 2$. Consequently, from Lemma 4, the only possibly non-vanishing terms of E_{2j} are those arising from the polynomials $S_{\alpha_1, \cdots, \alpha_t}$, $T'_{\alpha_1, \cdots, \alpha_t}$ with $\alpha_1 = \cdots = \alpha_s = 2$, and $\alpha_{s+1} = \cdots = \alpha_t = 1$, so that $2s + t - s = 2j$ or $t = 2j - s$, $s = 0, 1, \cdots, j$. For such values of $\alpha_1, \cdots, \alpha_t$ we have

$$C^k_{\alpha_1, \cdots, \alpha_t} = \frac{(2j)!}{2^s}.$$

Furthermore, as shown below, in developing $S_{\alpha_1,\cdots,\alpha_t}$ by means of Lemma 4 the coefficient of S_2^j is

$$(4.2) \qquad (-1)^{j-s}\frac{(2j-2s)!}{2^{j-s}(j-s)}.$$

DEMONSTRATION OF (4.2): If $s = j$, then it follows from Lemma 4 that the coefficient of S_2^j is 1. If $s < j$, we use Lemma 4, and noting that $S_1 = 0$, we obtain

$$(4.3) \qquad S_{\alpha_1,\cdots,\alpha_t} = -S_{\alpha_1+\alpha_t,\alpha_2,\cdots,\alpha_{t-1}} - \cdots - S_{\alpha_1,\cdots,\alpha_{t-2},\alpha_{t-1}+\alpha_t},$$

where, since $\alpha_t = 1$, we have $\alpha_1 + \alpha_t = \alpha_2 + \alpha_t = \cdots = 1$, $\alpha_s + \alpha_t = 3$, and $\alpha_{s+1} + \alpha_t = \cdots = \alpha_{t-1} + \alpha_t = 2$. Consequently of the $t - 1$ terms of the above evaluation of $S_{\alpha_1,\cdots,\alpha_t}$, exactly s will have α's > 2 and $t - s - 1$ will be of the same form as $S_{\alpha_1,\cdots,\alpha_t}$ except that instead of s of the α's being 2 we have $s + 1$ of the α's equal 2. For each such s we repeat the process obtaining

$$S_{\alpha_1,\cdots,\alpha_t} = (-1)^{(t-s)/2}(t-s-1)(t-s-3)\cdots 3\cdot 1 \quad S_{\underbrace{2,\cdots,2}_{j}}$$

$$+ \text{ terms which have } h < j.$$

Consequently (4.2) provides the coefficient of S_2^j in $S_{\alpha_1,\cdots,\alpha_t}$. Since the other terms of $S_{\alpha_1,\cdots,\alpha_t}$ have $h < j$, they lead to terms of E_{2j} that vanish in the limit.

Furthermore, by Lemma 3, $T_{\alpha_1,\cdots,\alpha_t} = T'_{\alpha_1,\cdots,\alpha_t}s!(t-s)!$ and the only term of $T_{\alpha_1,\cdots,\alpha_t}$ for which $g = t$ is

$$T_2'^{s}T_1'^{t-s} = n^{(t-s)/2}A_n^{t-s}.$$

The other terms of $T_{\alpha_1,\cdots,\alpha_t}$ will lead to terms of E_{2j} that vanish in the limit since $g < t$. Consequently, eliminating terms known to tend to zero as n and N become infinite, we see that $E_{2j} - f(n, N)$ tends to zero as n and N become infinite, where

$$f(n, N) = \sum_{s=0}^{j}\frac{(2j)!}{2^s}(-1)^{j-s}\frac{(2j-2s)!\,N^j n^{j-s}A_n^{2j-2s}}{2^{j-s}s!(2j-2s)!\,N^{(2j-s)}(1-nA_n^2/N)^j}.$$

Now as n and N become infinite with $n < N$, we see that

$$\lim f(n, N) = \lim \frac{(2j)!}{2^j}\sum_{s=0}^{j}(-1)^{j-s}\frac{1}{s!(j-s)!}(nA_n^2/N)^{j-s}/(1-nA_n^2/N)^j$$

$$= \frac{(2j)!}{2^j j!},$$

i.e.,

$$\lim E_{2j} = \frac{(2j)!}{2^j\cdot j!}.$$

To complete the proof it is only necessary to note that the normal distribution is completely determined by its moments.[6]

---

[6] See for example, M. G. Kendall, *The Advanced Theory of Statistics*, Vol. I, London, Charles Griffin and Company, page 110.

Since Theorem 2 is a generalization of the Theorem of Wald and Wolfowitz, it is possible to generalize slightly all the applications they make of their theorem. The statements of these generalizations are omitted

The application of Theorem 2 that led to this paper is the following: Suppose that $c_{jn} = n^{-1/2}$. Then the sequence C_n satisfies W and $A_n = 1$. Consequently we have proved

COROLLARY 1. *If the sequence U_N satisfies the condition W and if $\bar{x}_n$ is the arithmetic mean of a sample of n elements selected at random without replacement from U_N, then, for all a,*

$$\lim P\left\{\frac{n^{1/2}(\bar{x}_n - \bar{x}_N)}{\mu_{2N}^{1/2}(1 - m/n)^{\frac{1}{2}}} < a\right\} = \left(\frac{1}{2\pi}\right)\int_{-\infty}^{a} e^{-x^2/2}\, dx,$$

provided that $\epsilon > 0$ exists such that $n/N < 1 - \epsilon$, if n and N are sufficiently large.

Now the sequence of U_N will certainly satisfy W if U_N has the same moments for all values of N, or if the moments of U_N tend to fixed values as N increases, or if the universe U_N is a random sample of a universe having these properties. Consequently Theorem 1 and its corollaries will be valid for many applications, among them being the case studied by F. N. David[7] when U_N has the same multinomial distribution for each value of N.

The condition W is immediately satisfied for large classes of changing universes. For example, if the elements of all U_N are uniformly bounded and

$$\lim \mu_{2N} \neq 0,$$

then the condition W is satisfied. As an illustration, consider the case where U_N contains Np_N elements having the value one and $N(1 - p_N)$ elements having the value zero. Then

$$\mu_{2N} = p_N(1 - p_N),$$

and

$$\mu_{rN} = \frac{1}{N}\sum_{r=1}^{Np_N} (1 - p_N)^r + \sum_{r=Np_N+1}^{N} (-p_N)^r,$$

$$= p_N(1 - p_N)^r + (-1)^r(1 - p_N)p_N^r.$$

Hence

$$\frac{\mu_{rN}}{\mu_{2N}^{r/2}} = \frac{(1 - p_N)^{r/2}}{p_N^{r/2-1}} + (-1)^r \frac{p_N^{r/2}}{(1 - p_N)^{r/2-1}},$$

so that condition W will be satisfied if $\epsilon > 0$ exists such that $\epsilon < p_N < 1 - \epsilon$ for all sufficiently large N.

Hence the limiting distribution of Z_n will be normal no matter how the proportions p_N change provided only that the universe U_N does not come to consist essentially only of zeros or only of ones.

[7] Op. cit.

Various multivariate extensions of Theorem 2 are immediate. For example:

THEOREM 3. *Suppose that the elements of* U_N *are vectors of two components,*[8] $(x_{\nu N1}, x_{\nu N2})$, *and that the condition W is satisfied by the sequences* C_n, U_{N1}, *and* U_{N2} *where* U_{Nh}, $h = 1, 2$, *contains the elements* $x_{\nu Nh}$.

Let

$$z_{nh} = \sum_j c_{jn} x'_{jnh},$$

and let

$$Z_{nh} = \frac{z_{nh} - Ez_{nh}}{\sigma_{z_{nh}}},$$

where the random variables x'_{jnh} *are defined as were* x'_{jn}.

Let

$$\rho_N = \frac{\Sigma_\nu (x_{\nu N1} - \bar{x}_{N1})(x_{\nu N2} - \bar{x}_{N2})}{(\mu_{2N1} \cdot \mu_{2N2})^{1/2}},$$

and suppose that $\lim \rho_N$ *exists and is equal to* ρ *where* $\rho > -1 + \epsilon$. *Then, the limiting distribution of* Z_{n1} *and* Z_{n2} *is bivariate normal with means* 0, *variances* 1, *and correlation coefficient* ρ.

PROOF: To prove Theorem 3 we shall show that any linear function $t_1 Z_{n1} + t_2 Z_{n2}$ will be normally distributed in the limit if t_1 and t_2 are not both zero. It will then follow[9] that the theorem is true.

If we define $\hat{U}_N$ to be the sequence whose elements are

$$\hat{x}_{\nu N} = \frac{t_1(x_{\nu N1} - \bar{x}_{N1})}{\mu_{2N1}^{1/2}} + \frac{t_2(x_{\nu N2} - \bar{x}_{N2})}{\mu_{2N2}^{1/2}},$$

then the arithmetic mean of $\hat{U}_N$ is zero. Let

$$\hat{z}_n = \sum_j c_{jn} \hat{x}'_{jN},$$

and let

$$\hat{Z}_n = \frac{\hat{z}_n - E\hat{z}_n}{\sigma_{\hat{z}_n}}.$$

Then, it is readily verified that

$$\hat{Z}_n = \frac{t_1 Z_{n1} + t_2 Z_{n2}}{\sigma_{t_1 Z_{n1} + t_2 Z_{n2}}}.$$

---

[8] The generalization holds for any finite number of components but, to simplify the discussion, is stated for two components only. The method used is due to H. Cramér, *Random Variables and Probability Distributions*, Cambridge University Press, London, 1937, p. 105

[9] H Cramér, *Random Variables and Probability Distributions*, Cambridge University Press, London, 1937, p 105

Consequently, to prove that $t_1Z_{n1} + t_2Z_{n2}$ has a normal limiting distribution, we need to verify that the sequence U_N satisfies the condition W if U_{N1} and U_{N2} do

The moments of $\hat{U}_N$ are

$$\hat{\mu}_{rN} = \frac{1}{N} \sum_{\nu} \hat{x}_{\nu N}^{r},$$

so that

$$\hat{\mu}_{2N} = t_1^2 + t_2^2 + 2t_1t_2\rho_N ,$$

where ρ_N has the usual form of the correlation coefficient. Furthermore, using the binomial expansion, we have

$$\hat{\mu}_{rN} = \sum_{\alpha=0}^{r} C_\alpha^r \frac{t_1 t_2^{r-\alpha} \mu_{\alpha, r-\alpha N}}{\mu_{2N1}^{\alpha/2} \mu_{2N2}^{(r-\alpha)/2}}, \tag{4.4}$$

where

$$\mu_{\alpha, r-\alpha N} = \frac{1}{N} \sum_{\nu} (x_{\nu N1} - \bar{x}_{N1})^\alpha (x_{\nu N2} - \bar{x}_{N2})^{r-\alpha}$$

Then, by the Cauchy-Schwarz inequality we have

$$\left| \sum_{\nu} (x_{\nu N1} - \bar{x}_{N1})^\alpha (x_{\nu N2} - \bar{x}_{N2})^{r-\alpha} \right| \leq \left[\sum_{\nu} (x_{\nu N1} - \bar{x}_{N1})^{2\alpha} \cdot \sum_{\nu} (x_{\nu N2} - \bar{x}_{N2})^{2r-2\alpha} \right]^{\frac{1}{2}},$$

so that

$$| \mu_{\alpha, r-\alpha N} | \leq \mu_{2\alpha, N1}^{1/2} \mu_{2r-2\alpha, N2}^{1/2},$$

and using condition W for U_{N1} and U_{N2}, we have

$$\mu_{2\alpha, N1} \leq \mu_{2N1}^{\alpha} \lambda(N), \qquad \mu_{2r-2\alpha N2} \leq \mu_{2N2}^{r-\alpha} \lambda(N).$$

Hence, substituting in (4.4) we see that

$$\sup | \mu_{rN} | < \infty .$$

Hence the sequence U_N satisfies the condition W for all t_1 and t_2, and Theorem 3 is proved.

From Theorem 3, it then follows that the theorems on the limiting distributions of moments, product moments and functions of moments[10] are valid for sampling from finite universes, at random without replacement.

[10] The most important of these theorems are given in H Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1940, sections 28 2–28 4, pp. 364–367

A NON-PARAMETRIC TEST OF INDEPENDENCE[1]

By Wassily Hoeffding

Institute of Statistics, University of North Carolina

1. Summary. A test is proposed for the independence of two random variables with continuous distribution function (d.f.). The test is consistent with respect to the class Ω'' of d.f.'s with continuous joint and marginal probability densities (p.d.). The test statistic D depends only on the rank order of the observations. The mean and variance of D are given and $\sqrt{n}(D - ED)$ is shown to have a normal limiting distribution for any parent distribution. In the case of independence this limiting distribution is degenerate, and nD has a non-normal limiting distribution whose characteristic function and cumulants are given. The exact distribution of D in the case of independence for samples of size $n = 5, 6, 7$ is tabulated. In the Appendix it is shown that there do not exist tests of independence based on ranks which are unbiased on any significance level with respect to the class Ω''. It is also shown that if the parent distribution belongs to Ω'' and for some $n \geq 5$ the probabilities of the $n!$ rank permutations are equal, the random variables are independent.

2. Introduction. In a non-parametric test of a statistical hypothesis we do not make any assumptions about the functional form of the population distribution. A general theory of non-parametric tests is not yet developed, and a satisfactory definition of "best" non-parametric tests does not seem to be available Desirable properties of a "good" non-parametric test are unbiasedness and consistency. A test of a hypothesis H_0 is said to be consistent with respect to a specified class of admissible hypotheses if the probability of accepting H_0 tends to zero with increasing sample size whenever a hypothesis $\neq H_0$ of this class is true.

In this paper we consider the problem of testing the independence of two random variables X, Y on the basis of a random sample of size n. In all that follows the d.f. $F(x, y)$ of (X, Y) is assumed to be continuous. We will denote by Ω' the class of continuous d.f.'s $F(x, y)$ and by Ω'' the class of d.f.'s having continuous joint and marginal p.d.'s,

$$f(x, y) = \partial^2 F(x, y)/\partial x \, \partial y, \; f_1(x) = \int f(x, y) \, dy, \; f_2(y) = \int f(x, y) \, dx.$$

The hypothesis H_0 to be tested is that $F(x, y)$ is of the form

$$F(x, y) = F(x, \infty) F(\infty, y).$$

Several tests of this hypothesis have been proposed. Among them those deserve particular attention which depend only on the rank order of the obser-

[1] Research under a contract with the Office of Naval Research for development of multivariate statistical theory.

vations. They will be referred to as rank tests The critical region of a rank test of independence with respect to the class Ω' is similar to the sample space, the rank tests share this property with other tests obtained by the method of randomization (cf. Scheffé [1]). A characteristic feature of a rank test is that it remains invariant under order preserving transformations of X or Y.

Rank tests of independence have been studied by Hotelling and Pabst [2], Kendall [3] and Wolfowitz [4] While nothing is yet known about the power of the last test, the author [5] has shown that the two former tests are asymptotically biased for certain alternatives belonging to Ω' By a slight modification of the examples given in [5] it can be shown that these tests are asymptotically biased even with respect to the class Ω''.

In the Appendix it is shown that there do not exist rank tests of independence which are unbiased on any level of significance with respect to the classes Ω' or Ω''. It will appear from this paper that there do exist rank tests of independence which are consistent, and hence asymptotically unbiased, at least with respect to Ω''.

3. The Functional $\Delta(F)$. Given a random sample from a population with a d.f. belonging to a class Ω, we want to test the hypothesis H_0 that F is in a subclass ω of Ω. It is easy to construct a consistent test of H_0 if there exist (a) a functional $\theta(F)$ defined for every F in Ω and such that $\theta(F) = 0$ if and only if $F \epsilon \omega$; and (b) a consistent estimate of $\theta(F)$. There are many ways of devising by this method consistent tests of independence. The particular test described in the sequel has been chosen mainly for its relative simplicity.

If $F(x, y)$ is a bivariate d.f., let

$$D(x, y) = F(x, y) - F(x, \infty)F(\infty, y)$$

and

$$\Delta = \Delta(F) = \int D^2(x, y)\, dF(x, y). \tag{3.1}$$

Here and in the following, when no domain of integration is indicated, the (Lebesgue-Stieltjes) integral is extended over the entire space (here R_2).

The random variables X, Y with the d.f. $F(x, y)$ are independent if and only if $D(x, y) \equiv 0$.

THEOREM 3.1. *When $F(x, y)$ belongs to Ω'', $\Delta(F) = 0$ if and only if $D(x, y) \equiv 0$.*

PROOF. Evidently $D(x, y) \equiv 0$ implies $\Delta(F) = 0$.

Now suppose that $D(x, y) \not\equiv 0$. Since $F(x, y)$ is in Ω'', the function $d(x, y) = f(x, y) - f_1(x)f_2(y)$ is continuous. We have

$$D(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} d(u, v)\, du\, dv.$$

$D(x, y) \not\equiv 0$ implies $d(x, y) \not\equiv 0$, and since

$$\iint d(x, y)\, dx\, dy = 0,$$

there exists a rectangle Q in R_2 such that $d(x, y) > 0$ if (x, y) is in Q. Hence $D(x, y) \neq 0$ almost everywhere in Q, and $f(x, y) > 0$ in Q. Thus

$$\Delta(F) \geq \iint_Q D^2(x, y)\, f(x, y)\, dx\, dy > 0.$$

This completes the proof.

If $F(x, y)$ is discontinuous, we can have $\Delta(F) = 0$ and $D(x, y) \neq 0$. This is, for instance, the case for the distribution

$$P\{X = 0, Y = 1\} = P\{X = 1, Y = 0\} = \tfrac{1}{2}.$$

The question remains open whether $\Delta = 0$ implies $D(x, y) \equiv 0$ if $F(x, y)$ is continuous or absolutely continuous

In Section 7 it will be shown that

$$0 \leq \Delta \leq \tfrac{1}{30}$$

The upper bound $\frac{1}{30}$ is attained when $F(x, y)$ is the (continuous) d.f. of a random variable (X, Y) such that X has any continuous d.f and $Y = X$ (or, more generally, Y is a monotone function of X).

Let

$$C(u) = \begin{cases} 1 & \text{if } u \geqq 0, \\ 0 & \text{if } u < 0, \end{cases}$$

(3.2) $$\psi(x_1, x_2, x_3) = C(x_1 - x_2) - C(x_1 - x_3),$$

$$\phi(x_1, y_1; \cdots, x_5, y_5) = \tfrac{1}{4}\psi(x_1, x_2, x_3)\psi(x_1, x_4, x_5)\psi(y_1, y_2, y_3)\psi(y_1, y_4, y_5).$$

Then we can write

(3.3) $$\Delta = \int \cdots \int \phi(x_1, y_1; \cdots; x_5, y_5)\, dF(x_1, y_1) \cdots dF(x_5, y_5).$$

4. The Statistic D. Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be a random sample from a population with the d.f. $F(x, y)$, $n \geq 5$, and let

(4.1) $$D = D_n = \frac{1}{n(n-1) \cdots (n-4)} \Sigma'' \phi(X_{\alpha_1}, Y_{\alpha_1}; \cdots; X_{\alpha_5}, Y_{\alpha_5}),$$

where Σ'' denotes summation over all α such that

$$\alpha_i = 1, \cdots, n; \qquad \alpha_i \neq \alpha_j \text{ if } i \neq j, \qquad (i, j = 1, \cdots, 5).$$

Since the number of terms in Σ'' is $n(n-1) \cdots (n-4)$, we have by (3.3),

(4.2) $$ED = \Delta.$$

Since in the case of independence $ED = 0$, D can assume both positive and negative values. It will be seen in Section 7 that $-\frac{1}{60} \leq D_n \leq \frac{1}{30}$, the upper bound $\frac{1}{30}$ being attained for every n, while the minimum of D_n apparently increases with n.

The random variable D as defined by (4.1) belongs to the class of U-statistics considered by the author [5]. The following properties of D follow immediately from the results of that paper:

I *Let*

$$\Phi(x_1, y_1; \cdots; x_5, y_5) = D_5 = \frac{1}{5!}\Sigma'' \phi(x_{\alpha_1}, y_{\alpha_1}, \cdots, x_{\alpha_5}, y_{\alpha_5}),$$

$$\Phi_k(x_1, y_1; \cdots; x_k, y_k) = \int \cdots \int \Phi(x_1, y_1; \cdots, x_k, y_k; x_{k+1}, y_{k+1}; \cdots x_5, y_5)$$
$$dF(x_{k+1}, y_{k+1}) \cdots dF(x_5, y_5), \ (k = 1, \cdots, 5),$$

$$\zeta_k = \int \cdots \int \{\Phi_k(x_1, y_1, \cdots; x_k, y_k) - \Delta\}^2 \, dF(x_1, y_1) \cdots dF(x_k, y_k).$$

Then the variance of D_n *is*

$$\text{var } D_n = \binom{n}{5}^{-1} \sum_{k=1}^{5} \binom{5}{k}\binom{n-5}{5-k}\zeta_k \tag{4.3}$$

We have

$$25\,\zeta_1 \leq n \text{ var } D_n \leq 5\,\zeta_5 .$$

n var D_n is a decreasing function of n, and

$$\lim_{n\to\infty} n \text{ var } D_n = 25\,\zeta_1 . \tag{4.4}$$

II. By Theorem 7 1, [5], *the random variable* $\sqrt{n}(D_n - \Delta)$ *has a normal limiting distribution with mean zero and variance* $25\,\zeta_1$.

It will be seen in section 6 that in the case of independence $\zeta_1 = 0$, so that the normal limiting distribution of $\sqrt{n}D_n$ is a degenerate one. In this case nD_n has a non-normal limiting distribution (See section 8).

5. Computation of D. From (4.1) and (3.2) we get after reduction

$$D = \frac{A - 2(n-2)B + (n-2)(n-3)C}{n(n-1)(n-2)(n-3)(n-4)}, \tag{5.1}$$

where

$$A = \sum_{\alpha=1}^{n} a_\alpha(a_\alpha - 1)\, b_\alpha(b_\alpha - 1),$$

$$B = \sum_{\alpha=1}^{n} (a_\alpha - 1)(b_\alpha - 1)\, c_\alpha , \tag{5.2}$$

$$C = \sum_{\alpha=1}^{n} c_\alpha(c_\alpha - 1),$$

and

$$a_\alpha = \sum_{\beta=1}^{n} C(X_\alpha - X_\beta) - 1, \qquad b_\alpha = \sum_{\beta=1}^{n} C(Y_\alpha - Y_\beta) - 1,$$

$$c_\alpha = \sum_{\beta=1}^{n} C(X_\alpha - X_\beta)C(Y_\alpha - Y_\beta) - 1.$$

$a_\alpha + 1$ and $b_\alpha + 1$ are the ranks of X_α and Y_α, respectively. c_α is the number of sample members (X_β, Y_β) for which both $X_\beta < X_\alpha$ and $Y_\beta < Y_\alpha$. (Since $F(x, y)$ is continuous we may assume that $X_\alpha \neq X_\beta$ and $Y_\alpha \neq Y_\beta$ if $\alpha \neq \beta$.)

Thus, to compute D for a given sample we have to determine the numbers a_α, b_α, c_α for each sample member, calculate A, B, C from (5.2) and insert them in (5.1).

6. The variance of D in the case of independence. Since $F(x, y)$ is assumed to be continuous, so are $F(x, \infty)$ and $F(\infty, y)$. The inequalities $x_1 < x_2$ and $F(x_1, \infty) < F(x_2, \infty)$ are then equivalent unless $F(x_1, \infty) = F(x_2, \infty)$. The same is true of $y_1 < y_2$ and $F(\infty, y_1) < F(\infty, y_2)$. This shows that the function ϕ, (3.2), does not change its value if x_i, y_i is replaced by $F(x_i, \infty)$, $F(\infty, y_i)$, except perhaps on a set of zero probability. Hence Δ and D are invariant under the transformation

$$u = F(x, \infty), \qquad v = F(\infty, y); \qquad U = F(X, \infty), \qquad V = F(\infty, Y).$$

In the case of independence we have $F(x, y) = uv$, and

$$\zeta_k = \int_0^1 \cdots \int_0^1 \{\Phi_k'(u_1, v_1; \cdots; u_k, v_k)\}^2 \, du_1 \, dv_1 \cdots du_k \, dv_k,$$

where Φ_k' is defined as Φ_k, with x_i, y_i and $F(x_i, y_i)$ replaced by u_i, v_i and $u_i v_i$ respectively. On evaluation of these definite integrals we get

$$\zeta_1 = 0, \qquad 200 \cdot 30^2 \zeta_2 = \tfrac{2}{9}, \qquad 600 \cdot 30^2 \zeta_3 = \tfrac{14}{9},$$
$$600 \cdot 30^2 \zeta_4 = \tfrac{104}{9}, \qquad 120 \cdot 30^2 \zeta_5 = 12.$$

On inserting these values in (4.3) we obtain

$$\text{(6.1)} \qquad \operatorname{var}(30D) = \frac{2(n^2 + 5n - 32)}{9n(n-1)(n-3)(n-4)}.$$

Another way to determine the coefficients ζ_k in the case of independence is to compute $\operatorname{var} D_n$ for $n = 5, 6, 7$ from the exact distributions given in section 7, and $\lim_{n\to\infty} n^2 \operatorname{var} D_n$ from the asymptotic distribution of nD_n (section 8).

7. The exact distribution of D in the case of independence for $n = 5, 6, 7$. Let $S = \{(x_1, y_1), \cdots, (x_n, y_n)\}$ be a sample from a population with a continuous d.f. We may confine ourselves to samples with $x_i \neq x_j$ and $y_i \neq y_j$ if $i \neq j$. Let $(x_1', y_{\beta_1}'), \cdots, (x_n', y_{\beta_n}')$ be a rearrangement of $(x_1, y_1), \cdots, (x_n, y_n)$ such that $x_1' < x_2' < \cdots < x_n'$ and $y_1' < y_2' < \cdots < y_n'$. The permutation $\Pi = (\beta_1, \cdots, \beta_n)$ of $(1, \cdots, n)$ will be referred to as the ranking of the sample S.

D_n depends only on the ranking of the sample. We shall express this by writing $D_n = D_n(\Pi) = D_n(\beta_1, \cdots, \beta_n)$. If $(\beta'_{\alpha_1}, \cdots, \beta'_{\alpha_m})$ is a permutation of $m(< n)$ of the integers $1, \cdots, n$ such that $\beta'_1 < \beta'_2 < \cdots < \beta'_m$, $D_m(\beta'_{\alpha_1}, \cdots, \beta'_{\alpha_m})$ is defined to be equal to $D_m(\alpha_1, \cdots, \alpha_m)$. Replacing in (4.1) (X_α, Y_α) by (α, β_α) we find

$$D_n(\beta_1, \cdots, \beta_n) = \binom{n}{5}^{-1} \Sigma' D_5(\beta_{\alpha_1}, \cdots, \beta_{\alpha_5}), \tag{7.1}$$

where Σ' stands for summation over all α such that $1 \le \alpha_1 < \alpha_2 < \cdots < \alpha_5 \le n$

Denoting by $\Pi^{(i)}$ the permutation obtained from $\Pi = (\beta_1, \cdots, \beta_n)$ by omitting β_i, we have the recursion formula

$$nD_n(\Pi) = (n - 5) \sum_{i=1}^{n} D_{n-1}(\Pi^{(i)}). \tag{7.2}$$

From (4.1) and (3.2) we obtain

$$60D_5(\beta_1, \cdots, \beta_5) = \psi(\beta_3, \beta_1, \beta_4)\psi(\beta_3, \beta_2, \beta_5) + \psi(\beta_3, \beta_1, \beta_5)\psi(\beta_3, \beta_2, \beta_4)$$

or

$$60D_5(\beta_1, \cdots, \beta_5) = \begin{cases} 0 & \text{if } \beta_3 \ne 3; \\ 2 & \text{if } \beta_3 = 3 \text{ and } \beta_1, \beta_2 < 3 \quad \text{or } \beta_1, \beta_2 > 3; \\ -1 & \text{if } \beta_3 = 3 \text{ and } \beta_1 < 3, \beta_2 > 3 \text{ or } \beta_1 > 3, \beta_2 < 3. \end{cases} \tag{7.3}$$

We have

$$\begin{aligned} D_n(\beta_1, \cdots, \beta_n) &= D_n(\beta_2, \beta_1, \beta_3, \cdots, \beta_n) \\ &= D_n(\beta_1, \cdots, \beta_{n-2}, \beta_n, \beta_{n-1}) = D_n(\beta_n, \beta_{n-1}, \cdots, \beta_1) \end{aligned} \tag{7.4}$$

For $n = 5$ this follows from (7.3) and for general n from (7.1).

Also, by the symmetry of D_n with respect to x and y, D_n does not change its value if in the permutation $(\beta_1, \cdots, \beta_n)$ the numbers 1, 2 or $n - 1, n$ are interchanged or the permutation is replaced by its inverse

In the case of independence all $n!$ rankings have the same probability $1/n!$. To find the distribution of D_n we have to determine the number of rankings giving rise to particular values of D_n.

If $n = 5$ there are $5! = 120$ rankings. Owing to (7.4) we need consider only those with $\beta_1 < \beta_2$, $\beta_4 < \beta_5$, $\beta_1 < \beta_4$. Their number is $\frac{120}{8} = 15$. Among them those with $\beta_3 \ne 3$ yield $D_5 = 0$; this leaves only the three permutations

$$(1, 2, 3, 4, 5), \qquad (1, 4, 3, 2, 5), \qquad (1, 5, 3, 2, 4).$$

By (7.3) the respective values of $60D_5$ are $2, -1, -1$ Thus we have

$$P\{60D_5 = 2\} = \tfrac{1}{15}, \qquad P\{60D_5 = -1\} = \tfrac{2}{15},$$
$$P\{60D_5 = 0\} = \tfrac{12}{15}.$$

The distribution of D_6, D_7, $\cdots$ can be obtained in a similar way using the relations (7.1) to (7.4). The distribution of D_n for $n = 5, 6, 7$ is given in Table I.

From (7.3) and (7.1) it follows that $-\frac{1}{60} \leq D_n \leq \frac{1}{30}$ for $n = 5, 6, \cdots$. The upper bound $\frac{1}{30}$ is attained for $\Pi = (1, 2, \cdots, n)$ and every n. To judge by the cases $n = 5, 6, 7$, the minimum of D_n apparently increases with n. From $ED_n = \Delta$ it also follows that $\Delta \leq \frac{1}{30}$.

8. The Asymptotic Distribution of nD_n in the Case of Independence.

THEOREM 8.1. *If $F(x, y) = F(x, \infty)F(\infty, y)$ and $F(x, \infty)$ and $F(\infty, y)$ are continuous, the random variable $nD_n + \frac{1}{36}$ has a limiting distribution whose characteristic function* (c.f.) *is*

$$g(t) = \prod_{k=1}^{\infty} \left(1 - \frac{2it}{k^2 \pi^4}\right)^{-\frac{1}{2}\tau(k)} \tag{8.1}$$

where $\tau(k)$ is the number of divisors of k.

Note that $\tau(k)$ is the number of divisors of k including 1 and k. Thus $\tau(1) = 1$, $\tau(2) = 2$, $\tau(3) = 2$, $\tau(4) = 3$, $\cdots$.

The author has not been able to bring the d.f. corresponding to the c.f. $g(t)$ into a form suitable for numerical computation. Thus Theorem 8.1 may be considered as a preliminary result. For this reason only a brief indication of the proof is given here.

If $(X_1, Y_1), \cdots, (X_n, Y_n)$ is a random sample from a population with d.f. $F(x, \infty)F(\infty, y)$, let $nS_n(x, y)$ be the number of sample members (X_i, Y_i) such that $X_i \leq x$, $Y_i \leq y$. $S_n(x, y)$ is a d.f. depending on the random sample. If we put $F(x, y) = S_n(x, y)$ in $\Delta(F)$ as defined by (3.3), we get

$$\Delta(S_n) = \frac{1}{n^5} \sum_{\alpha_1=1}^{n} \cdots \sum_{\alpha_5=1}^{n} \phi(X_{\alpha_1}, Y_{\alpha_1}; \cdots; X_{\alpha_5}, X_{\alpha_5}).$$

It is easy to prove that if $n\{\Delta(S_n) - E\Delta(S_n)\}$ has a limiting distribution, it is the same as that of nD_n.

Now it can be shown that $n\Delta(S_n)$ has a limiting distribution with the c.f. (8.1). This can be done either analogously to Smirnoff's [6] derivation of the limiting distribution of the goodness of fit statistic ω_n^2, or applying von Mises' [7] general results on the asymptotic distribution of a differentiable statistical function. Though the latter paper deals only with univariate distributions, its results can be extended to the multivariate case.

By expanding $\log g(t)$ in powers of it we obtain for the j-th cumulant κ_j

$$\kappa_j = \frac{2^{5j-3}(j-1)!}{[(2j)!]^2} B_{2j-1}^2,$$

where B_{2j-1} are Bernoulli's numbers,

$$B_1 = \tfrac{1}{6}, \qquad B_3 = \tfrac{1}{30}, \qquad B_5 = \tfrac{1}{42}, \qquad B_7 = \tfrac{1}{30}, \cdots.$$

In particular, $\kappa_1 = \frac{1}{36}$, and since $ED_n = 0$, the limiting distribution of $n\Delta(S_n)$ is that of $nD_n + \frac{1}{36}$.

9. The D-test of Independence. Given a random sample from a bivariate population with continuous d.f., a test for independence can now be carried out as follows:

If $\alpha(0 < \alpha < 1)$ is the desired level of significance, let ρ_n be the smallest number satisfying the inequality

$$P\{D_n > \rho_n \mid F \epsilon \omega\} \leq \alpha,$$

where ω is the class of d.f.'s of the form $F(x, \infty)F(\infty, y)$.

Compute D_n as shown in section 5. Reject the hypothesis H_0 of independence if and only if $D_n > \rho_n$.

For $n = 5, 6, 7$ the numbers ρ_n can be obtained from Table I.

From Tchebychef's inequality and (6.1) we have

$$P\left\{ 30D_n > \sqrt{\frac{2(n^2 + 5n - 32)}{9n(n-1)(n-3)(n-4)\alpha}} \right\} \leq \alpha.$$

Hence

$$30\rho_n \leq \sqrt{\frac{2(n^2 + 5n - 32)}{9n(n-1)(n-3)(n-4)\alpha}}.$$

It follows that $\rho_n = O(n^{-1})$.

If $\Delta > 0$, we have $\Delta - \rho_n > 0$ for sufficiently large n. Then

$$P\{D_n > \rho_n\} \geq P\{\, |D_n - \Delta| \leq \Delta - \rho_n\} \geq 1 - (\text{var } D_n)/(\Delta - \rho_n)^2$$

By (4.4) the right hand side tends to 1.

This, together with Theorem 3.1, shows that the D-test is consistent with respect to the class Ω''.

Since $P\{D_n \leq 0\}$ tends to 0 if $\Delta > 0$, it is safe not to reject H_0 whenever $D_n \leq 0$. An inspection of Table I shows that at least for small n this will happen in more than one-half of the cases if H_0 is true.

10. Concluding Remarks. It would be interesting to compare the power of the D-test with that of other tests with respect to particular alternatives, for instance with the product moment correlation test when the population is normal with correlation ρ. A preliminary investigation seems to indicate that for small values of $|\rho|$ and $n \to \infty$ the power efficiency of the D-test as compared with the product moment correlation test is rather low. This result may not be conclusive for values of n which are of practical interest. On the other hand, it may be expected that a test which is consistent with respect to a large class of alternatives will have a lower power with regard to a sub-class of alternatives than a test which has optimum properties with respect to this particular sub-class. These considerations suggest the problem of selecting from a given class of non-para-

metric tests (such as those consistent with respect to Ω'') a test which is most powerful with respect to certain parametric alternatives (such as normal distributions).

TABLE I

The distribution of D_n in the case of independence for $n = 5, 6, 7$.

$n = 5$

x	$15P\{60D_5 = x\}$	$P\{60D_5 \geq x\}$
-1	2	1.0000
0	12	0.8667
2	1	0.0667

$n = 6$

x	$90P\{180D_6 = x\}$	$P\{180D_6 \geq x\}$
-2	4	1.0000
-1	28	0.9556
0	36	0.6444
1	16	0.2444
2	1	0.0667
3	4	0.0556
6	1	0.0111

$n = 7$

x	$630P\{1260D_7 = x\}$	$P\{1260D_7 \geq x\}$
-11	8	1.0000
-8	32	0.9873
-7	32	0.9365
-6	8	0.8857
-5	28	0.8730
-4	88	0.8286
-3	64	0.6889
-2	56	0.5873
-1	8	0.4984
0	88	0.4857
2	77	0.3460
3	24	0.2238
4	4	0.1857
6	56	0.1794
8	8	0.0905
9	4	0.0778
12	24	0.0714
14	2	0.0333
18	12	0.0302
24	2	0.0111
30	4	0.0079
42	1	0.0016

APPENDIX

A. Equiprobable rankings and independence. Let $\Pi_{n\nu}$, $(\nu = 1, 2, \cdots, n!)$ be the $n!$ possible rankings of samples of size n from a bivariate population with continuous d.f. $F(x, y)$ (cf. section 7).

If $F(x, y) = F(x, \infty)F(\infty, y)$ we have

$$P\{\Pi_{n\nu}\} = 1/n! \qquad (\nu = 1, \cdots, n!) \tag{A1}$$

for every n.

Does (A1) for some particular n imply independence? This is not true for $n = 2$. In this case (A1) is equivalent to $P\{(1, 2)\} = \frac{1}{2}$. If the distribution has a p.d. $f(x, y)$, we have

$$P\{(1, 2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)\, du\, dv + \int_{x}^{\infty}\int_{y}^{\infty} f(u, v)\, du\, dv\right] f(x, y)\, dx\, dy,$$

which equals $\frac{1}{2}$ whenever $f(x, y) = f(-x, y)$. However, we have the following theorem:

THEOREM. *If $F(x, y)$ is in Ω'' and* (A1) *holds for some $n \geq 5$, then*

$$F(x, y) = F(x, \infty)F(\infty, y). \tag{A2}$$

PROOF. (4.2) can be written in the form

$$\sum_{\nu=1}^{n!} D_n(\Pi_{n\nu}P)\{\Pi_{n\nu}\} = \Delta. \tag{A3}$$

If (A1) holds, the left hand side of (A3) has the same value as when (A2) is true. But in the latter case we have $\Delta = 0$. Hence (A1) implies $\Delta = 0$ By Theorem 3 1 this is sufficient for (A2). The proof is complete.

B. Non-existence of unbiased rank tests of independence.

THEOREM. *There do not exist rank tests of independence which are unbiased on any significance level with respect to the classes Ω' or Ω''.*

PROOF: Let $\Pi_{n\nu}$ have the meaning of Appendix A. Any critical region of a rank test of independence is a set $S_m = \{\Pi_{n\nu_1}, \cdots, \Pi_{n\nu_m}\}$ of m rankings In the case of independence $P(S_m) = P\{\Pi_{n\nu} \in S_m\} = m/n!$ We may confine ourselves to significance levels $m/n!$, $m = 1, 2, \cdots, n! - 1$. To prove the theorem it is sufficient to show that for every $n = 2, 3, \cdots$, for some $m(1 \leq m \leq n! - 1)$ and every S_m there exists a d.f. F in Ω'' such that

$$P(S_m \mid F) < m/n!.$$

We shall prove the slightly more general proposition that this holds for

$$m = 1, 2, 3.$$

Let the bivariate distribution A_n be such that the probability mass is distributed uniformly on the $n - 1$ segments

$$\frac{k-1}{n-1} < x \leq \frac{k}{n-1}, \qquad y - x = \frac{n-2k}{n-1}, \qquad (k = 1, 2, \cdots, n-1), \tag{B1}$$

and is zero in any region not containing a part of these segments.

Let B_n be the distribution which is uniform on the $n - 1$ segments

$$\frac{k-1}{n-1} < x \leq \frac{k}{n-1}, \qquad x + y = \frac{2k-1}{n-1}, \qquad (k = 1, 2, \cdots, n-1), \tag{B2}$$

and zero elsewhere.

The d.f.'s of both A_n and B_n are continuous, with

$$F(x, \infty) = F(\infty, x) = x \qquad (0 \leq x \leq 1).$$

Since the probability of (X, Y) lying on any one of the segments (B1) or (B2) is $1/(n-1)$, the probabilities $P(\Pi/A_n)$ and $P(\Pi/B_n)$ are easily obtained in terms of the multinominal distribution with $n - 1$ equal probabilities. In particular, we have

$$\text{(B3)} \qquad P(1, 2, \cdots, n \mid A_2) = 1; \qquad P(n, n-1, \cdots, 1 \mid B_2) = 1,$$

$$\text{(B4)} \qquad \begin{aligned} P(1, 2, \cdots, n \mid A_n) = P(n, n-1, \cdots, 1 \mid B_n) &= (n-1)\left(\frac{1}{n-1}\right)^n \\ &= \left(\frac{1}{n-1}\right)^{n-1}, \\ P(n, n-1, \cdots, 1 \mid A_n) = P(1, 2, \cdots, n \mid B_n) &= 0. \end{aligned}$$

In general, if Π_n is any permutation of $1, \cdots, n$, we have either $P(\Pi_n \mid A_n) = 0$ or $P(\Pi_n \mid B_n) = 0$. For any Π_n with $P(\Pi_n \mid A_n) \neq 0$ contains at least one "run up" of 2 or more numbers (a sequence of consecutive numbers $i, i+1, \cdots, i+k$) which is not preceded by smaller numbers or followed by larger numbers. On the other hand, if a Π_n' with $P(\Pi_n' \mid B_n) \neq 0$ contains a "run up", it is either preceded by smaller numbers or followed by larger numbers. Hence if $P(\Pi_n \mid A_n) \neq 0$, then $P(\Pi_n \mid B_n) = 0$. Similarly, $P(\Pi_n \mid B_n) \neq 0$ implies $P(\Pi_n \mid A_n) = 0$.

From (B3) it follows that for any set S_m of m rankings which does not include $(1, 2, \cdots, n)$ or $(n, n-1, \cdots, 1)$ we have either $P(S_m \mid A_2) = 0$ or $P(S_m \mid B_2) = 0$. Hence we need only consider critical regions containing both $(1, 2, \cdots, n)$ and $(n, n-1, \cdots, 1)$. For $m = 1$ there are no such regions. For $m = 2$ there is just one. But from (B4) it follows that for $n > 2$,

$$P(1, 2, \cdots, n \mid A_n) + P(n, n-1, \cdots, 1 \mid A_n) = \left(\frac{1}{n-1}\right)^{n-1} < \frac{2}{n}\left(\frac{1}{n-1}\right)^{n-2} < \frac{2}{n!}.$$

Finally, if Π_n is any permutation other than $(1, 2, \cdots, n)$ or $(n, n-1, \cdots, 1)$, we have, by the preceding arguments, either for A_n or for B_n,

$$P(1, 2, \cdots, n) + P(n, n-1, \cdots, 1) + P(\Pi_n) = \left(\frac{1}{n-1}\right)^{n-1} < \frac{3}{n!}.$$

This completes the proof for d.f.'s in Ω'. To prove the theorem for d.f.'s in Ω'' we can replace the distributions A_n and B_n by distributions A_n' and B_n' having continuous joint and marginal densities and such that the probabilities $P(\Pi \mid A_n')$ and $P(\Pi \mid B_n')$ differ as little as we please from $P(\Pi \mid A_n)$ and $P(\Pi \mid B_n)$, respectively. For instance, A_2' can be defined by the continuous density

$$
\begin{aligned}
f(x, y) &= K(\epsilon - y + x) && \text{if } 0 \le y - x \le \epsilon, \quad x \le 1 - \epsilon, y \ge \epsilon; \\
&= K(\epsilon - x + y) && \text{if } -\epsilon \le y - x \le 0, \quad x \ge \epsilon, \quad y \le 1 - \epsilon; \\
&= K(x + y - \epsilon) && \text{if } x + y \ge \epsilon, \quad x \le \epsilon, \quad y \le \epsilon; \\
&= K(2 - \epsilon - x - y) && \text{if } x + y \le 2 - \epsilon, x \ge 1 - \epsilon, y \ge 1 - \epsilon; \\
&= 0 && \text{elsewhere,}
\end{aligned}
$$

where $K = 3/(3\epsilon^2 - 4\epsilon^3)$ and $0 < \epsilon \le \frac{1}{2}$. If ϵ is taken sufficiently small, the distribution satisfies the requirements. The details are left to the reader.

The proof also shows the non-existence of an unbiased rank test of independence for $n = 2$ and any level of significance (for we need consider only one level, $\frac{1}{2}$). It also can be shown that for $n = 3$, any $m = 1, 2, \cdots, 5$ and any S_m the inequality $P(S_m) < m/3!$ holds for at least one of the distributions A_2, A_3, B_2, B_3. The question remains open whether there exist rank tests of independence which are unbiased for some sample sizes n and some significance levels $m/n!$.

REFERENCES

[1] H. Scheffé, "Statistical inference in the non-parametric case," *Annals of Math. Stat*, Vol. 14 (1943), pp 305–332.

[2] H Hotelling and M. R. Pabst, "Rank correlation and tests of significance involving no assumptions of normality," *Annals of Math Stat.*, Vol 7 (1936), pp. 29–43.

[3] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, Vol. 30 (1938), pp. 81–93.

[4] J. Wolfowitz, "Additive partition functions and a class of statistical hypotheses," *Annals of Math. Stat.*, Vol 13 (1942), pp. 247–279.

[5] W. Hoeffding, "A class of statistics with asymptotically normal distribution," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 293–325.

[6] N. V. Smirnoff, "On the distribution of Mises' ω^2-criterion," (Russian, with French summary), *Matematicheskii Sbornik*, Nov. Ser., Vol. 2 (1937), pp. 973–993.

[7] R. von Mises, "On the asymptotic distribution of differentiable statistical functions," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 309–348.

ON PREDICTION IN STATIONARY TIME SERIES

By Herman O. A. Wold

Uppsala University

Summary. In time series analysis there are two lines of approach, here called the *functional* and the *stochastic*. In the former case, the given time series is interpreted as a mathematical function, in the latter case as a random specimen out of a universe of mathematical functions. The close relation between the two approaches is in section 2 shown to amount to a genuine isomorphism Considering the problem of prediction from this viewpoint, the author gives in sections 3–4 the functional equivalence of his earlier theorem on the decomposition of a stationary stochastic process with a discrete time parameter (see [9], theorem 7). In section 5 the decomposition theorem is applied to the problem of linear prediction. Finally in section 6 a few comments are made. Since various aspects of the isomorphism in question are known, this paper might be regarded as essentially expository.

1. Introductory. Let the sequence

$$\cdots, x_{t-1}, x_t, x_{t+1} \cdots \tag{1}$$

be an empirical time series such that no clear trend is present in the average level, in the variance or in any other structural properties of the series which we might choose to consider. Such series are usually called *stationary* as distinct from *evolutive*, terms which of course are somewhat loose when referring to empirical data. We shall consider two approaches in the theoretical analysis of stationary series. It is convenient to allow x_t to be complex; the conjugate complex of x_t is denoted $\bar{x}_t$.

In the functional approach, the sequence (1) is regarded as forming an infinite sequence, say $\{x_t\}$, where t runs from $-\infty$ to $+\infty$. To define stationarity, let us for any infinite sequence $\{z_t\}$ write

$$M[z_t] = \lim \frac{1}{t_2 - t_1 + 1} \sum_{t=t_1}^{t_2} z_t \qquad (t_1 \to -\infty,\ t_2 \to +\infty). \tag{2}$$

The limit $M[z_t]$, which will be called "the average of z_t", is clearly independent of t. It is also seen that a necessary and sufficient condition for $M[z_t]$ to exist is that the same average should be obtained when t_1 is kept fixed while $t_2 \to +\infty$, and when t_2 is kept fixed while $t_1 \to -\infty$. The stationarity of the sequence (1) may now be brought out by assumptions of the type that the averages $M[x_t]$ and $M[x_t \cdot \bar{x}_{t+k}]$ exist, say

$$M[x_t] = m, \qquad M[x_t \cdot \bar{x}_{t+k}] = r_k \qquad (k = 0, \pm 1, \pm 2, \cdots). \tag{3}$$

In the stochastic (or probabilistic) approach, we introduce an infinite sequence of random variables, say

$$\cdots, \xi_{t-1}, \xi_t, \xi_{t+1}, \cdots \qquad (-\infty < t < +\infty), \tag{4}$$

or briefly $\{\xi_t\}$. The sequence $\{\xi_t\}$ may be regarded as the generalization of the notion of multi-dimensional variable, say $[\xi_1, \cdots, \xi_n]$, to an infinite number of components ξ_t. According to a basic theorem by A. Kolmogoroff (see e g. [9], §11), the probability distribution of the sequence $\{\xi_t\}$ may be defined by specifying for any finite set of variables, say $[\xi_{t_1}, \cdots, \xi_{t_n}]$, its multi-dimensional distribution function, say

(5) $$F(u_1, \cdots, u_n ; t_1, \cdots, t_n) = \text{Prob}\,(\xi_{t_1} \leq u_1, \cdots, \xi_{t_n} \leq u_n).$$

The sequence $\{\xi_t\}$ thus defined is said to constitute a *stochastic process*. As is sufficient for our purpose, we confine ourselves to the case when the time parameter t is restricted to discrete values, $t = 0, \pm 1, \pm 2, \cdots$.

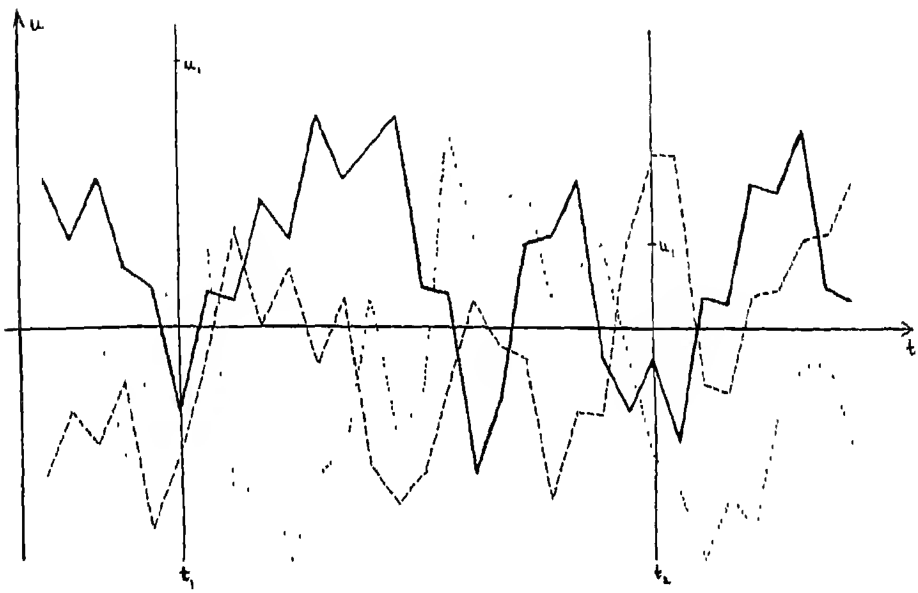


FIG. 1

Now in the stochastic approach, the empirical time series (1) is regarded as a sample specimen, a *realization*, of the stochastic process $\{\xi_t\}$, just as a point $[x_1, \cdots, x_n]$ in an n-dimensional space may be regarded as a sample specimen of a multidimensional variable $[\xi_1, \cdots, \xi_n]$. In line with this interpretation, the process $\{\xi_t\}$ may be regarded as a universe of individual realizations such as (1) (see the graph). Taking out a realization at random from this universe, we shall have the probability,

$$F(u_1 ; t_1) = \text{Prob}\,(\xi_{t_1} \leq u_1),$$

that the value taken on by the realization at the time point t_1 will be $\leq u_1$; similarly,

$$F(u_1, u_2 ; t_1, t_2) = \text{Prob}\,(\xi_{t_1} \leq u_1, \xi_{t_2} \leq u_2),$$

is the joint probability that the values taken on by the realization at t_1 and t_2 will be $\leq u_1$ and $\leq u_2$ respectively.

Any expectation referring to the variables (4) may be expressed in terms of the distribution functions (5), for instance

$$E[\xi_t] = \int_{-\infty}^{\infty} u\, d_u F(u; t), \qquad E[\xi_{t_1} \cdot \xi_{t_2}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u \cdot v\, d^2_{u,v} F(u, v; t_1, t_2).$$

Again interpreting in terms of the universe of realizations, $E[\xi_t]$, say, is the average, over this universe, of the value taken by the realizations at the time point t.

The above definition of a stochastic process (4) being perfectly general, we have to impose special assumptions if we wish to take into account particular properties of the given time series (1). Thus stationarity of the process (4) may be defined by assuming that any probability of the type (5) will remain the same if $t_1, \cdots, t_n$ is replaced by $t_1 + t, \cdots, t_n + t$, where t is arbitrary. Alternatively, and more generally, the stationarity of the sequence (1) may be brought out in this approach by assuming that the expectations

$$E[\xi_t] = \mu, \qquad E[\xi_t \cdot \xi_{t+k}] = \rho_k$$

exist and are independent of t.

2. The functional and stochastic approaches are closely related as to problems and results. A typical example is that r_k and ρ_k as defined above allow the representations[1]

$$(6) \qquad r_k = \int_{-\pi}^{\pi} e^{ik\lambda}\, dF(\lambda), \quad \rho_k = \int_{-\pi}^{\pi} e^{ik\lambda}\, d\Phi(\lambda), \qquad (k = 0, \pm 1, \pm 2, \cdots),$$

where $F(\lambda)$ and $\Phi(\lambda)$ are real, bounded and never decreasing functions. We shall now show that the parallelism between the two approaches amounts to a mathematical isomorphism. On the one hand, we recall that A. Kolmogoroff [3], [4] has introduced and studied the notion of a stationary sequence in Hilbert space,—let such a sequence be denoted $\{X_t\}$—, and shown that a stationary stochastic process $\{\xi_t\}$ forms a particular realization of this general, abstract $\{X_t\}$. On the other hand the following elementary lemma shows that another realization of $\{X_t\}$ may be formed on the basis of a stationary sequence $\{x_t\}$ such as (1).

LEMMA. *Let $\{x_t\}$ be a sequence of type* (1) *which satisfies the conditions* (3) *but is arbitrary in other respects. We write*

$$(7) \qquad \{\mathbf{x}_t\} = \cdots, \mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1}, \cdots,$$

where $\mathbf{x}_t = \{x_t\}$, and $\mathbf{x}_{t+k}$ is obtained from $\mathbf{x}_t$ by replacing x_t by x_{t+k} for every t.

[1] As to r_k, see N. Wiener [8], who treats the case of a continuous time parameter t. As to ρ_k, see H. Wold [9], p. 66, and A. Kolmogoroff [4], p. 5

For the elements x_t, *let multiplication by a real or complex constant and addition be defined by*

$$ax_t = \{ax_t\}, \qquad x_t + y_t = \{x_t + y_t\},$$

and let R *be the class formed by all elements of the type*

$$c_{-n}x_{t-n} + c_{-n+1}x_{t-n+1} + \cdots + c_0x_t + \cdots + c_nx_{t-n},$$

where n *and* $c_{-n}, \cdots, c_n$ *are arbitrary. Let the inner product* (x_t, y_t) *of two elements* $x_t = \{x_t\}$, $y_t = \{y_t\}$ *in* R *be defined by*

$$(x_t, y_t) = M[x_t \cdot \bar{y}_t],$$

and let R' *be the closure of* R

Then R' *is a space the dimension of which is denumerable or finite. In the former case,* R' *satisfies the conditions of a Hilbert space* H, *in the latter case it can be extended to a Hilbert space* H. *In any case, the relations*

$$(8) \qquad Ux_t = x_{t+1}, \qquad -\infty < t < +\infty,$$

define a unitary transformation U *in* H.

The first statement of the theorem is obvious. It is also easily verified that R' satisfies the conditions A–C of an abstract Hilbert space as defined by B. v Sz. Nagy [7]. If R' is of finite dimension, a suitable extension will make R' satisfy the conditions A–E of a Hilbert space as defined by M. H. Stone [6]. The transformation U is clearly unitary, it is also plain that the definition (8) of U extends to the whole of H.

Now since both (4) and (7) are particular realizations of a stationary sequence $\{X_t\}$ in Hilbert space, any theorem on such a sequence $\{X_t\}$ will give, as immediate corollaries, similar theorems on a stationary sequence $\{x_t\}$ of type (1) and on a stationary stochastic process $\{\xi_t\}$. Generally speaking, the former corollary will involve averages of one or more functional sequences $\{x_t\}$, $\{y_t\}$, $\cdot$ over time t, while the latter will involve averages, for fixed t, over the realizations of one or more stochastic processes $\{\xi_t\}$, $\{y_t\}$, $\cdots$.

Let us consider the following problem of prediction in the light of the isomorphism established: Suppose the data (1) are known up to $t - 1$, say for $t - 1, t - 2, \cdots, t - n$, what can then be said about x_t, or, more generally, about x_{t+k}? One approach to the problem is to apply harmonic analysis to the given data, and to extrapolate the function obtained up to the time point $t + k$ Another approach, the one which we shall consider, is to approximate x_{t+k} directly in terms of the given data Confining ourselves to linear prediction, and making use of n observations, the prediction formula will then be

$$(9) \qquad \text{pred. } x_{t+k} = a_0^{(n,k)} + a_1^{(n,k)}x_{t-1} + a_2^{(n,k)}x_{t-2} + \cdots + a_n^{(n,k)}x_{t-n}.$$

The error of prediction, also called the residual, is denoted

$$(10) \qquad y_{t+k}^{(n,k)} = x_{t+k} - \text{pred } x_{t+k}.$$

Considering first the functional approach, we apply formula (9) for all t, thus obtaining the residuals

$$\cdots, \quad y_{t-1}^{(n,k)}, \quad y_t^{(n,k)}, \quad y_{t+1}^{(n,k)}, \quad \cdots.$$

In this approach we are led to regard the residual variance, i.e.

$$M[\,|\,y_t^{(n,k)}\,|^2], \tag{11}$$

as a total measure of the accuracy of the prediction. If we follow the stochastic approach, on the other hand, the formula (9) is applied, for fixed t, to all realizations $\{x_t\}$ of the process $\{\xi_t\}$. In this case, the variance expectation,

$$E[|\,y_t^{(n,k)}\,|^2], \tag{12}$$

is regarded as a total measure of the accuracy of the prediction. The prediction coefficients $a_i^{(n,k)}$ are determined by minimizing the expressions (11) and (12), respectively.[2] It needs no further comment that the two lines of approach in prediction theory will, thanks to the isomorphism indicated, lead to parallel results.

In a study of stationary stochastic processes, the author has earlier found a decomposition theorem which has a direct bearing on the prediction problem (see [9], theorem 7). The main purpose of the present note is to develop the corresponding decomposition for a functional sequence of the type (1). Two theorems on this line are given in sections 3–4. The proofs are briefly indicated; for further details, the reader is referred to my treatment on the stationary process [9]. In section 5, the decomposition is applied to the prediction problem. A few comments follow in section 6.

3. Auto-regression analysis of stationary time series. Let $\{x_t\}$ be an infinite sequence (1) such that the conditions (3) are fulfilled. By (9)–(10), the residuals $y_t^{(n,0)}$ will be well-defined for every n and t. According to elementary properties of least square residuals, we have

$$M[y_t^{(n,0)}] = 0; \qquad M\,[y_t^{(n,0)}\cdot \bar{x}_{t-k}] = 0 \text{ for } k = 1, 2, \cdots, n. \tag{13}$$

Since the minimum variance cannot increase if we replace n by $n+1$, we further have

$$M[\,|\,x_t\,|^2] \geq M[\,|\,y_t^{(n,0)}\,|^2] \geq M\,[\,|\,y_t^{(n+1,0)}\,|^2] \geq 0.$$

Making $n \to \infty$, we infer that there is a constant d^2 such that

$$\lim_{n\to\infty} M[|\,y_t^{(n,0)}\,|^2] = d^2 \geq 0.$$

[2] For real sequences $\{x_t\}$ and $\{\xi_t\}$, this minimization is, of course, nothing else than the method of least squares.

Making use of the Gram-Schmidt orthogonalization procedure, it is further possible to show that there exists a sequence $\{y_t\}$ such that

$$\lim_{n\to\infty} M[|y_t^{(n,0)} - y_t|^2] = 0.$$

In the usual terminology, the sequence $\{y_t\}$ is the *limit in the mean* of the sequence $\{y_t^{(n,0)}\}$,

$$\text{(14)} \qquad \underset{n\to\infty}{\text{l.i.m.}} (\cdots, y_{t-1}^{(n,0)}, y_t^{(n,0)}, y_{t+1}^{(n,0)}, \cdots) = \cdots y_{t-1}, y_t, y_{t+1}, \cdots.$$

We may remark that (14) does not necessarily imply that $y_t^{(n)}$ will for a fixed t have y_t for an ordinary limit We also note that the limiting sequence $\{y_t\}$ is not uniquely determined, for instance, the relation (14) remains valid if a finite number of the elements y_t are modified

As is easily shown, we have

$$\text{(15)} \qquad \lim_{n\to\infty} M[|y_t^{(n,0)}|^2] = M[|y_t|^2] = M[y_t \cdot \bar{x}_t] = d^2 \geq 0,$$

and [cf. (13)]

$$\text{(16)} \qquad M[y_t \bar{x}_{t-k}] = 0, \qquad k = 1, 2, \cdots.$$

Moreover, the sequence $\{y_t\}$ is non-autocorrelated, i.e.

$$\text{(17)} \qquad M[y_t \bar{y}_{t+k}] = 0, \qquad k = \pm 1, \pm 2, \cdots.$$

In fact, observing that

$$M[y_t y_{t+k}] = \lim_{n\to\infty} M[y_t^{(n,0)} \cdot y_{t-k}^{(n,0)}] \qquad k = 1, 2, \cdots,$$

and supposing that (17) is not true, we would have

$$\text{(18)} \qquad | M[y_t^{(\nu,0)} \cdot \bar{y}_{t-k}^{(\nu,0)}] | > a > 0,$$

as ν runs through some sequence $n_1, n_2, \cdots$, such that $n_\nu \to \infty$. The relation (18), however, would imply

$$\text{(19)} \qquad M[\, | y_t^{(\nu,0)} - c\, y_{t-k}^{(\nu,0)} |^2] \leq d^2 (1 - \tfrac{1}{2}a^2)$$

for some sufficiently large ν and for some suitable c. Since $y_t^{(\nu,0)} - c\, y_{t-k}^{(\nu,0)}$ is a linear expression of the type appearing in the right hand member of (9), the relation (19) is incompatible with (15). Thus (18) is not possible and (17) must hold good.

Part of the above analysis is summed up in

THEOREM 1. *Given a time series $\{x_t\}$ which satisfies (3), let $\epsilon > 0$ be arbitrary. Then an integer n and a set of coefficients $a_i^{(n,0)}$ exist for which (9) defines a residual series $\{y_t^{(n,0)}\}$ such that*

$$M[y_t^{(n,0)}] = 0, \qquad | M[y_t^{(n,0)} \cdot \bar{y}_{t+k}^{(n,0)}] | < \epsilon \qquad k = \pm 1, \pm 2, \cdots.$$

4. A decomposition theorem. We shall first consider the special case where (15) gives

$$M[|y_t|^2] = d^2 = 0, \tag{20}$$

which is the same as

$$\underset{n\to\infty}{\text{l.i.m.}} (\cdots y_{t-1}^{(n,0)}, y_t^{(n,0)}, \cdots) = (\cdots, 0, 0, \cdots).$$

In this case we shall say that the sequence $\{x_t\}$ is deterministic,[3] the interpretation of this term being as follows: Given the sequence $\{x_t\}$ for all time points up to and including $t - 1$, we may, by the use of a finite number of the given values, predict x_{t+k} with any accuracy; i.e., with a residual error of arbitrarily small variance This can be shown by induction. In fact, suppose that we are able to predict each of $x_t, \cdots, x_{t+k-1}$ in such a way that the prediction error has a variance $< \epsilon$, where ϵ is arbitrarily prescribed. Letting $\delta > 0$ be arbitrary, we can then find a formula of type (9) which predicts x_{t+k} in terms of the exact values $x_{t+k-1}, x_{t+k-2}, \cdots$ and which gives a residual variance $\delta/(k + 1)$. Replacing here $x_{t+k-1}, \cdots, x_t$ by values so predicted that the residual variances are less than $\delta/(k + 1) | a_1^{(n,0)} |, \cdots, \delta/(k + 1) | a_k^{(n,0)} |$, it is seen that the total error of (9) will have a variance $< \delta$.

We proceed to the *general case*, $d^2 \geq 0$. According to the above analysis, y_t is that part of x_t which cannot be linearly predicted from the previous observations $x_{t-1}, x_{t-2}, \cdots$. In other words, each time point t brings in an unpredictable, random-like element y_t in the series $\{x_t\}$. Now while from (16) y_t is uncorrelated with the previous observations $x_{t-1}, x_{t-2}, \cdots$, it will in general be correlated with the future observations $x_{t+1}, x_{t+2}, \cdots$. Thus the unpredictable element y_t may be regarded as influencing the future development $x_{t+1}, x_{t+2}, \cdots$ of the series $\{x_t\}$. In order to examine this influence we proceed as follows.

We approximate x_t linearly in terms of $y_t, y_{t-1}, \cdots, y_{t-n}$, writing

$$x_t = b_0 y_t + b_1 y_{t-1} + \cdots + b_n y_{t-n} + u_t^{(n)} = z_t^{(n)} + u_t^{(n)}.$$

Determining the coefficients b_k by minimizing

$$M[|x_t - z_t^{(n)}|^2],$$

the coefficients b_k will thanks to (16)–(17) be independent of n. We obtain

$$b_0 = 1; \qquad b_k = M[x_t \cdot \bar{y}_{t-k}]/d^2, \qquad k = 1, 2, \cdots.$$

The sequence $\{z_t^{(n)}\}$ thus being determined for every n, it is further easily shown that $\{z_t^{(n)}\}$ converges in the mean, say to $\{z_t\}$,

$$\underset{n\to\infty}{\text{l.i.m.}} (\cdots, z_{t-1}^{(n)}, z_t^{(n)}, \cdots) = (\cdots, z_{t-1}, z_t, \cdots). \tag{21}$$

[3] The term is due to J Doob [1], in my study [9] I used the term *singular*.

We may thus write

$$z_t = y_t + b_1 y_{t-1} + b_2 y_{t-2} + \cdots,$$

where the sum converges in the mean. Finally, we write

(22) $$x_t = z_t + u_t,$$

which gives a decomposition of the series $\{x_t\}$ into two components $\{z_t\}$ and $\{u_t\}$.

In the decomposition (22) the component z_t is that part of x_t which is linearily built up by the unpredictable elements $\{y_t\}$ up to and including the time point t. From (17) we know that the sequence $\{y_t\}$ is non-autocorrelated. It can further be shown that the square modulus sum of the coefficients b_k is convergent,

$$\sum_{k=0}^{\infty} |b_k|^2 < \infty.$$

As to the component u_t, it can be shown that $\{u_t\}$ is deterministic. More precisely, we have

$$\underset{n\to\infty}{\text{l.i.m.}} \{u_t - (a_0^{(n,0)} + a_1^{(n,0)} u_{t-1} + \cdots + a_n^{(n,0)} u_{t-n})\} = \{0\}$$

where the $a_i^{(n,0)}$ are the same as the minimizing coefficients of (9). It can further be shown that u_t is uncorrelated with y_{t+k} and z_{t+k} for all k,

$$M[u_t \bar{y}_{t+k}] = M[u_t \bar{z}_{t+k}] = 0, \qquad (k = 0, \pm 1, \pm 2, \cdots).$$

Summing up the above results, we obtain

THEOREM 2 *Any time series $\{x_t\}$ which satisfies the conditions (3) allows the decomposition*

(23) $$\{x_t\} = \{z_t + u_t\},$$

with

$$\{z_t\} = \underset{n\to\infty}{\text{l.i.m.}} \{y_t + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_n y_{t-n}\},$$

where the series $\{y_t\}$, $\{z_t\}$ and $\{u_t\}$ have the following properties

A. *The elements y_t, z_t and u_t are obtained from $x_t, x_{t-1}, \cdots$ by the limit formulae* (14), (21) *and* (22).

B. *The series $\{y_t\}$ has zero mean,*

$$M[y_t] = 0,$$

is non-autocorrelated,

$$M[y_t \bar{y}_{t+k}] = 0, \qquad k = \pm 1, \pm 2, \cdots,$$

and is uncorrelated with $\{x_{t-1}\}$, $\{x_{t-2}\}$, $\cdots$,

$$M[y_t \cdot \bar{x}_{t-k}] = 0, \qquad k = 1, 2, \cdots.$$

C. *The series* $\{u_t\}$ *is uncorrelated with* $\{y_t\}$ *and* $\{z_t\}$,

$$M[u_t\bar{y}_{t+k}] = M[u_t\bar{z}_{t+k}] = 0, \qquad (k = 0, \pm 1, \pm 2, \cdots).$$

D. *The series* $\{u_t\}$ *is deterministic.*

5. Application to the problem of prediction. In section 1 we have considered the problem of predicting x_{t+k} linearly in terms of $x_{t-1}, x_{t-2}, \cdots$. Now it is seen that theorem 2 gives the following formula for predicting x_{t+k} with an error of minimal variance,

$$\text{pred. } x_{t+k} = u_{t+k} + b_{k+1}y_{t-1} + b_{k+2}y_{t-2} + \cdots .$$

In fact, by theorem 2, A and D, the right-hand member can be calculated with any prescribed accuracy from a finite set of observations $x_{t-1}, x_{t-2}, \cdots, x_{t-N}$, where N of course depends on the accuracy desired; on the other hand, the prediction error being

$$y_{t+k} + b_1y_{t+k-1} + \cdots + b_ky_t,$$

we infer from theorem 2 (B) that this error is of minimal variance,

$$M[\,|\,x_{t+k} - \text{pred } x_{t+k}\,|^2] = (1 + |\,b_1\,|^2 + \cdots + |\,b_k\,|^2)d^2$$

6. Comments. As mentioned in section 2, the above theorem 2 is the analogue of a theorem on the decomposition of a stationary stochastic process given by the author previously (see [9], theorem 7). The starting point is then to apply formula (9), not as above to the same sequence $\{x_t\}$ for varying t, but to all realizations $\{x_t\}$ of the process, holding t fixed. The close connection between the decomposition in the two approaches is further brought out by the following theorem.

THEOREM 3. *Given a stochastic process,*

$$\cdots, \xi(t-1), \xi(t), \xi(t+1), \cdots,$$

which is stationary in the sense of (5), *let* $\{x_t\}$ *be an individual realization of this process Then* $\{x_t\}$ *will with probability* 1 *allow the decomposition of theorem* 2.

In fact, according to the ergodic theorem of Birkhoff-Khintchine,[4] the averages (2) will exist with probability 1, and so theorem 3 follows from theorem 2. It should be observed that the coefficients b_k will in general vary from one realization to another.

The theory of the decomposition (23) has been carried further in a brilliant study by A. Kolmogoroff [3]. His analysis deals with the general case of a stationary sequence in a Hilbert space. Establishing a decomposition of type (23)

[4] See A. Kolmogoroff [2]. His proof refers to averages (2) of the special type where t_1 is held fixed while $t_2 \to \infty$. According to the stationarity, however, the average exists, and is the same, when t_2 is fixed and $t_1 \to -\infty$, and so the general average (2) will likewise exist.

for such sequences Kolmogoroff also shows that the decomposition is uniquely determined by properties corresponding to A–D. Making use of the powerful methods of spectral analysis of linear transformations in Hilbert space, Kolmogoroff further presents a highly developed theory of the decomposition.

As immediate corollaries of this general theory Kolmogoroff [4] obtains corresponding results for a stationary stochastic process $\{\xi_t\}$ such as (4). Now thanks to our lemma in section 2, similar theorems hold good for the functional sequence (1). These results include detailed theorems on the connection between the decomposition (23) and, on the other hand, the function $F(\lambda)$ which by (6) generates the coefficients r_k. For example, it turns out that $\{x_t\}$ is completely deterministic if the derivative $F'(\lambda)$ is constant over an interval of positive measure. An explicit formula for the coefficients b_k in terms of the function $F(\lambda)$ may also be obtained. For proofs and further results, we must refer to Kolmogoroff's papers [3]–[4].

The theory of the decomposition (23) has later been generalized in various directions. V. Zasuhin [11] and J. Doob [1] have shown that the decomposition applies to multi-dimensional stationary sequences. As shown by the present author [10], the decomposition may be employed for the analysis of linear equation systems with an infinite number of unknowns. This device makes use of the decomposition of non-stationary sequences, a generalization indicated also by M. Loève [5].

REFERENCES

[1] J. L. Doob, "The elementary Gaussian processes," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 229–282.

[2] A. Kolmogoroff, "Ein vereinfachter Beweis des Birkhoff-Khintchineschen Ergodensatzes," *Rec. Math (Sbornik) N.S.*, Vol. 2 (1937), pp 367–368

[3] A. Kolmogoroff, "Stationary sequences in Hilbert space (Russian)," *Bolletin Moskovskovo Gosudarstvenovo Universiteta, Matematika*, Vol. 2 (1941).

[4] A Kolmogoroff, "Interpolation und Extrapolation von stationaren zufalligen Folgen," *Bull. Acad. Sci. URSS, Sér Math*, Vol 5 (1941), pp 3–14.

[5] M Loève, "Fonctions aléatoires de second ordre," *Revue Sci.*, Vol. 84 (1946), pp. 195–206.

[6] M H. Stone, *Linear Transformations in Hilbert Space*, Amer. Math. Soc. Colloq Publ 15, New York, 1932.

[7] B v Sz Nagy, *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes*, Ergebn. Math. u. Grenzgeb., Vol. 5, Berlin, 1942.

[8] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, Vol. 55 (1933), pp. 117–258.

[9] H. Wold, *A Study in the Analysis of Stationary Time Series*, Dissertation (Stockholm), Uppsala, 1938.

[10] H. Wold, "On infinite, non-negative definite, Hermitian matrices, and corresponding linear equation systems," *Arkiv Mat. Astr. Fys.*, Vol. 29A (1943)

[11] V Zasuhin, "On the theory [illegible] stationary random processes," *C R. (Doklady) Akad. Sci* [illegible] (1941), pp 435–437.

GENERALIZATION TO N DIMENSIONS OF INEQUALITIES OF THE TCHEBYCHEFF TYPE

By Burton H. Camp

Wesleyan University

1. Summary. The Tchebycheff statistical inequality and its generalizations are further generalized so as to apply equally well to n-dimensional probability distributions. Comparisons may be made with other generalizations [1], [2] that have been developed recently for the two-dimensional case. The inequalities given in this paper are generally as close as the most favorable corresponding inequalities that exist for the one-dimensional case and in many simple cases they are closer than those that have been given heretofore for two dimensions. In a special case the upper bound of our inequality is actually attained. The theory contains also a less important generalization in one dimension.

2. Introduction. It is necessary to introduce a new kind of moment, to be called a "contour" moment, which is a generalization of the usual one-dimensional moment. If we consider first a simple two-dimensional frequency surface, $y = f(t_1, t_2)$, we may think of y as a function of a single variable, x, where x is the area of the contour on that surface at the y level. This function may be defined so that it is monotonic decreasing and has other simple characteristics. Then we define the rth contour moment as

$$\bar{\mu}_r = \int_0^\infty x^r y \, dx,$$

and then the generalization of the Tchebycheff-type inequalities follows easily. This theory can be applied equally well to almost any single-valued function of n variables which is limited and integrable in the sense of Lebesgue. Therefore the theory will be enunciated initially in a very general form. The reasons for the initial statements will be indicated only briefly because a detailed discussion of quite similar ideas has been given by this author in another paper [3], where he applied the same general principle to obtain generalizations of certain theorems in integration theory.

3. Preliminary theory. Let $f(t_1, \cdots, t_n)$ be a probability distribution with limited upper bound L and defined at all points of infinite n-space, which is to be denoted by T, dT being the Lebesgue measure of a differential element. We thus assume that: $0 \leqq f(t, \cdots, t_n) \leqq L$, f has a Lebesgue integral in T, and

$$\int_T f \, dT = 1.$$

Let Q_λ denote the set of points in T where $f > \lambda$, $(0 \leqq \lambda \leqq L)$, and let x_λ be the

measure of Q_λ, for Q_λ is known to be measurable Therefore $x_L = 0$, $x_0 \leqq \infty$, and for each λ there exists a unique Q_λ and therefore a unique x_λ. This means that x is a single-valued function of λ and that it exists (or is positive infinite) for every value of λ in the interval $(0 \leqq \lambda \leqq L)$. If $\lambda' > \lambda$, $x_{\lambda'} \leqq x_\lambda$ This means that x is a monotonic decreasing function of λ. It need not be continuous; that is, it may be asymptotic to the line $\lambda = 0$, and it may have finite discontinuities or "jumps". Also there may be an enumerably infinite number of λ intervals in which x is constant. It follows that λ is a monotonic decreasing function of x in the interval $(0 \leqq x \leqq x_0 \leqq \infty)$, but it may not exist (in intervals where x has jumps), and it may be multiple valued (at points where x is constant) We now let $y(x) = \lambda_x$, except that: if λ is multiple valued at any point x we let y have the minimum value of λ at that point. Any other value would do equally well because the total measure of such points is zero and they can be left out of the integrals that follow. If λ does not exist in an x interval, we let y have in that interval the value which it has at the beginning of the interval. This is a λ point where x has a jump. We have thus defined y as a single valued monotonic decreasing function of x in the interval $(0 \leqq x \leqq x_0 \leqq \infty)$ and $0 \leqq y \leqq L$. It follows from Lebesgue's theory that:

$$\int_0^{x_\lambda} y(x)\,dx = \int_{Q_\lambda} f\,dT, \; (0 < \lambda \leqq L), \quad \int_0^{x_0} y(x)\,dx = \int_T f\,dT = 1.$$

Finally we restrict our function f so that there shall be at most a finite number of points x where λ is multiple valued (intervals of λ over which x is constant), and hence the number of discontinuities of y will be finite. This restriction may not be necessary but it is convenient and not embarrassing in applications.

4. Contour moments. The rth contour moment is denoted by $\hat{\mu}_r$. The contour standard deviation is denoted by $\hat{\sigma}$ We define

$$\hat{\mu}_r = \int_0^{x_0} x^r y\,dx.$$

It follows that $\mu_0 = 1$, and that

$$\hat{\mu}_2 = \hat{\sigma}^2 = \int_0^{x_0} x^2 y\,dx.$$

We shall also let $\hat{\alpha}_{2r} = \hat{\mu}_{2r}/\hat{\sigma}^{2r}$ We now assume that r is either zero or a positive integer, but in much of what follows this assumption is not necessary.

EXAMPLE 1. Let $f(t_1, t_2) = (2\pi)^{-1}e^{-(t_1^2+t_2^2)/2}$. The equation, $f(t_1, t_2) = \lambda$, defines a circular contour whose area is $x = \pi(t_1^2 + t_2^2) = -2\pi \log 2\pi\lambda$. Hence $y = \lambda = (2\pi)^{-1}e^{-x/2\pi}$, and

$$\hat{\mu}_r = \int_0^\infty x^r y\,dx = (2\pi)^r r!, \; \hat{\sigma}^2 = 8\pi^2, \; \hat{\alpha}_{2r} = (2r)!/2^r.$$

5. Contour moments and one-dimensional moments. If $n = 1$ and if $f(t_1) = f(-t_1)$, then

$$\hat{\mu}_{2r} = \int_0^{x_0} x^{2r} y\,dx = 2\int_0^{(x_0/2)} (2t)^{2r} f(t)\,dt = \mu_{2r} \cdot 2^{2r},$$

where μ_{2r} is an ordinary moment. Hence also $\hat{\sigma} = 2\sigma$, $\hat{\alpha}_{2r} = \hat{\mu}_{2r}/\hat{\sigma}^{2r} = \mu_{2r}\cdot 2^{2r}/\sigma^{2r}\cdot 2^{2r} = \alpha_{2r}$. It is to be noticed that, although $\hat{\alpha}_{2r} = \alpha_{2r}$, $\hat{\mu}_{2r} \neq \mu_{2r}$. One could alter the definition so that these two moments would be equal by inserting into the definition of contour moments the factor 2^n, using $x/2^n$ in place of x, but this would introduce a slight complication for a doubtful advantage. Although it would seem to be desirable to define the even contour moments $\hat{\mu}_{2r}$ so that they would become the ordinary moments μ_{2r} in the symmetrical one-dimensional case, such a definition would not make the two corresponding odd moments equal, and it would not make the two even moments equal in the non-symmetrical one-dimensional case. So it seems better not to introduce this factor 2^n, but to take note of the relationships that hold in the one-dimensional case.

THEOREM. *Let*

$$P_\delta = \int_{Q_\lambda} f\, dT,$$

where λ *is such that* $x_\lambda = \delta\hat{\sigma}$. *Then*

$$1 - P_\delta \leqq \hat{\alpha}_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}.$$

COROLLARY 1. *In particular* $1 - P_\delta \leqq \hat{\alpha}_{2r}/\delta^{2r}$.

COROLLARY 2. *If* $r = 1$, $1 - P_\delta \leqq 4/9\delta^2$. This theorem and these two corollaries are minor generalizations even of the corresponding one-dimensional inequalities, for it is no longer assumed that the probability distribution $f(t)$ has but one mode.

PROOF OF THEOREM. Let $g(x) = y(x)$ if $0 \leqq x \leqq x_0 \leqq \infty$, let $g(x) = y(-x)$ if $-\infty \leqq -x_0 \leqq x \leqq 0$, and let $g(x) = 0$ elsewhere in $(-\infty, \infty)$. Then $g(x)$ has all the properties explicitly required of $f(x)$ in a former paper by this author [4] in which this theorem was proved for the one-dimensional case. That is: $g(x)$ is a frequency function whose mean is zero, and

$$\int_{-\infty}^{\infty} g(x)\, dx = 2, \quad \text{and} \quad \int_{\delta\sigma}^{\infty} g(x)\, dx$$

is the probability that $|x| > \delta\sigma$; $g(x)$ is a monotonic decreasing function of $|x|$ for all values of x; and is symmetrical with respect to the central ordinate. Therefore, transforming the symbols of that paper to our present notation, we have

$$\int_{\delta\sigma}^{\infty} g(x)\, dx \leqq \alpha_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r},$$

where

$$\sigma^2 = \int_0^{\infty} x^2 g\, dx = \int_0^{\infty} x^2 y\, dx = \hat{\sigma}^2.$$

Similarly $\mu_{2r} = \hat{\mu}_{2r}$, $\alpha_{2r} = \hat{\alpha}_{2r}$, and finally

$$1 - P_\delta = 1 - \int_0^{\delta\theta} y\,dx = 1 - \int_0^{\delta\sigma} g\,dx = \int_{\delta\sigma}^{\infty} g\,dx$$

$$\leqq \alpha_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r} = \hat{\alpha}_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}.$$

This proves the theorem except that there is one exceptional case that requires attention. In the proof of the theorem in the paper just referred to the author assumed that the function corresponding to our present $g(x)$ was continuous. At that time a "frequency" function was often thought of as determined by a smooth curve approximating a histogram and implied even the existence of derivatives, and so continuity was not added to the explicit requirement that the function be a "frequency" function, but this condition was explicitly introduced in the lemma on which the proof of the theorem was based, and so we do now have to consider separately the case where y, and hence g, may have a finite number of jumps. It is quite easy to handle this case as the limiting form of a continuous case. In that lemma it was also required that d^2Q/dt^2 should exist and be non-negative, which would imply that we now have to make the requirement that y (corresponding to dQ/dt) shall have a non-negative first derivative. On examination of the proof, however, it will be observed that this is not necessary, since y is monotonic decreasing and continuous. That is, in the lemma the only use made of the condition, $d^2Q/dt^2 \geqq 0$, was that the function $Q(t)$ should determine a curve which would be never concave down. But for this it is sufficient that dQ/dt be continuous and monotonic increasing, and these conditions are now satisfied by the function which plays the rôle of Q in the present discussion. This function will now be defined as

$$\int_x^{\infty} \gamma(x)\,dx.$$

Let $\gamma(x)$ be a continuous function defined as equal to $g(x)$ except in the neighborhood of the points of finite discontinuity. Near such points it is to be so defined that it shall have all the properties just required of $g(x)$, and in addition so that, for any prescribed $R > 1$ and $\epsilon > 0$,

$$\int_0^{\infty} x^{2r} \gamma(x)\,dx = \int_0^{\infty} x^{2r} g(x)\,dx + \eta_r, \qquad (1 \leqq r \leqq R),$$

$$\int_x^{\infty} \gamma(x)\,dx = \int_x^{\infty} g(x)\,dx + \eta,$$

where $|\,\eta,\ \eta_r\,| < \epsilon$. It is obvious that such a definition of γ may be made in many ways, and one of them is by making use of a linear function in the neighborhood of each point of discontinuity. Since $\gamma(x)$ now satisfies all the conditions of the author's earlier paper the corresponding inequality is true:

$$\left(\int_{\delta\sigma_1}^{\infty} \gamma\,dx\right)\left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}\left(\int_0^{\infty} x^2\gamma\,dx\right)^{r} \leqq \int_0^{\infty} x^{2r}\gamma\,dx,$$

where

$$\sigma_1^2 = \int_0^\infty x^2 y\, dx.$$

Hence

$$\left(\int_{k\sigma_1}^\infty g\, dx - \eta\right)\left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}(\hat{\sigma}^2 - \eta_1)^r \leqq \hat{\mu}_{2r} - \eta_r.$$

Let ϵ approach zero and we have, as desired:

$$1 - P_\delta \leqq \hat{\alpha}_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}.$$

Example 2. Let

$$f(t_1, \cdots, t_n) = A \exp\left\{-\frac{1}{2}\left(\frac{t_1^2}{\sigma_1^2} + \cdots + \frac{t_n^2}{\sigma_n^2}\right)\right\}, \qquad A = (2\pi)^{-n/2}(\sigma_1 \cdots \sigma_n)^{-1}.$$

This is a form into which the general correlation solid may be put by means of a linear transformation. Since P_δ is a ratio between two parts of such a solid and since this ratio is preserved under a linear transformation, the more general case may be transformed into this one, or even, as will appear shortly, into the simpler one where all the standard deviations are unity. If $f = \lambda$ the contour is the ellipsoid,

$$\frac{t_1^2}{\sigma_1^2} + \cdots + \frac{t_n^2}{\sigma_n^2} = -2 \log \frac{\lambda}{A}.$$

The volume of this ellipsoid is

$$x = h(-2 \log \lambda/A)^{n/2}, \qquad h = V_0 \sigma_1 \cdots \sigma_n,\ V_0 = \frac{2\pi^{n/2}}{n\Gamma(n/2)}.$$

Hence $y = Ae^{-\frac{1}{2}(x/h)^{2/n}}$, $\quad \hat{\mu}_r = \int_0^\infty x^r y\, dx$

$$= nAh^{r+1}2^{n/2(r+1)-1}\Gamma\left(\frac{nr+n}{2}\right) = \left(\frac{\pi^{n/2}2^{n/2+1}\sigma_1 \cdots \sigma_n}{n}\right)^r \frac{\Gamma\left(\frac{nr+n}{2}\right)}{[\Gamma(n/2)]^{r+1}}.$$

Putting $r = 2$ we obtain

$$\hat{\sigma}^2 = \frac{\pi^n 2^{n+2}(\sigma_1 \cdots \sigma_n)^2}{n^2}\frac{\Gamma(3n/2)}{[\Gamma(n/2)]^3},$$

and then

$$\hat{\alpha}_{2r} = \frac{\hat{\mu}_{2r}}{\hat{\sigma}^{2r}} = \frac{\Gamma\left(\frac{2rn+n}{2}\right)}{\Gamma(n/2)}\left[\frac{\Gamma(n/2)}{\Gamma(3n/2)}\right]^r.$$

Our inequality becomes: $1 - P_\delta \leq J$, where

$$J = \frac{\hat{\alpha}_{2r}}{\left(\hat{\delta} \cdot \frac{2r+1}{2r}\right)^{2r}}, \text{ or } 1, \text{ whichever is smaller.}$$

Typical numerical values of $\hat{\alpha}_{2r}$ and of J are given in Tables I and II.

TABLE I

Values of $\hat{\alpha}_{2r}$

n	$\hat{\alpha}_{2r}$	$\hat{\alpha}_2$	$\hat{\alpha}_4$	$\hat{\alpha}_6$
1	$1 \cdot 3 \cdots (2r-1)$	1	3	15
2	$(2r)!/2^r$	1	6	90
3	$3 \cdot 5 \cdot 7 \cdots (6r+1)/(3 \cdot 5 \cdot 7)^r$	1	12.26	566
4	$(4r+1)!/(5!)^r$	1	25.20	3604

TABLE II

Values of J

δ	n	r	J
1	1	1	0.444
		2	1.000
1	2	1	0 444
		2	1.000
2	1	1	0 111
		2	0.077
		3	0 093
3	1	1	0.049
		2	0.015
		3	0.008
		4	0 006
		5	0.006
3	2	1	0.049
		2	0.030
		3	0.049
3	3	1	0.049
		2	0.062
		3	0.308

Let us now compare J with the true value of $(1 - P_\delta)$ in one of these cases, *viz.*, when $\delta = 3$ and $n = 3$. The true value is given by

$$1 - P_3 = 1 - A \int_0^{3\hat{\sigma}} e^{-\frac{1}{2}(x/h)^{2/3}}\, dx,$$

where now $\hat{\sigma} = 4\pi\sqrt{105}(\sigma_1\sigma_2\sigma_3)/3$, $h = 4\pi(\sigma_1\sigma_2\sigma_3)/3$. The integral may be evaluated by means of the transformation, $t = (x/h)^{1/3}$ and a table of the integral of $(2\pi)^{-1/2}e^{-t^2/2}(t^2 - 1)$. We obtain: $1 - P_3 = 0.0205$. This is the true value to be compared with the approximation, $J = 0.049$. The closeness of this approximation is similar to that which may be obtained for the normal law by using the corresponding inequalities for one dimension. To illustrate this we find from the usual tables that, if for the normal law $1 - P_\delta = 0.0205$, $\delta = 2.32$. Hence the corresponding inequality is (for $r = 2$): $1 - P_\delta \leqq 0.042$.

We shall now show that the upper bound of our inequality is actually attained in a special case. Let $f(t_1, \cdots, t_n) = 2^{-n}$ in the region $(-1 \leqq t_1, \cdots, t_n \leqq 1)$, and let $f = 0$ elsewhere. For this case we shall have $x = 0$ when $\lambda = 2^{-n}$, and $x = 2^n$ when $0 \leqq \lambda < 2^{-n}$. Therefore $y = 2^{-n}$ if $0 \leqq x < 2^n$, and $y = 0$ if $2^n \leqq x$. Hence $\hat{\sigma} = 2^n/\sqrt{3}$, $\mu_0 = 1$, and the true value of $(1 - P_\delta)$ is $1 - \delta/\sqrt{3}$; and when $\delta = 2/\sqrt{3}$, this true value is $1/3$. The appropriate inequality is: $1 - P_\delta \leqq 4/9\,\delta^2$, and when $\delta = 2/\sqrt{3}$ the right hand side of this inequality is also equal to $1/3$. These relationships are true for all values of n.

REFERENCES

[1] P. O. Berge, "A note on a form of Tchebycheff's inequality for two variables," *Biometrika*, Vol. 29 (1937), pp. 405–406.

[2] Z. W. Birnbaum, J. Raymond, and H. S. Zuckerman," A generalization of Tchebycheff's inequality to two dimensions," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 70–79.

[3] B. H. Camp, "A method of extending to multiple integrals properties of simple integrals," *Math. Ann.*, Vol. 75 (1914), pp. 274–289.

[4] B. H. Camp, "A new generalization of Tchebycheff's statistical inequality," *Amer. Math. Soc. Bull.*, Vol. 28 (1922), pp. 427–432. See also: "A note on Narumi's paper," *Biometrika*, Vol. 15 (1923), pp 421–423.

BOUNDARIES OF MINIMUM SIZE IN BINOMIAL SAMPLING

By R. L. Plackett

University of Liverpool

1. Introduction. Much attention has recently been concentrated on the problems arising when sampling a binomial population, since this is thought to form a suitable model for certain industrial and biological procedures. A general discussion of such procedures as applied in industry has been given by Barnard [2] and various particular cases have received detailed treatment by Burman [3] Stockman and Armitage [6], and Anscombe [1]. Unbiased estimation of the population parameter (the "fraction defective") has been investigated by Girshick, Mosteller and Savage [4] and Wolfowitz [7]. A paper by Haldane [5] is also relevant.

For such sampling procedures it is necessary to find the probabilities of accepting or rejecting material with a particular fraction defective; to calculate the average sample size, and to form an estimate of the fraction defective when sampling terminates. All three characteristics may be expressed in terms of quantities $N(x, y)$, defined in section 3, so that once these are known, the fundamental properties of the scheme are known.

Here we present a method for determining the $N(x, y)$, investigate the conditions under which it is valid; relate the method to the estimation problem; and exemplify its application. The schemes to which the method can successfully be applied are of a special type (to which the title refers) and include all inspection procedures with a finite upper limit to the sample size likely to be used in practice. Other schemes, when dissected in a manner similar to that used by Stockman and Armitage, can doubtless be formulated as an aggregate of the special types

2. Nomenclature. Our nomenclature differs in some respects from that of Girshick, Mosteller and Savage, although the same collection of terms is employed. References to their paper should therefore be followed by a comparison of the terminology

Taking a sample of one from a binomial population consists in observing either of two events, whose probabilities are p and $1 - p$ ($p \neq 0$ or 1). The results of successive samples of one can be represented by the path of a particle in a two-dimensional lattice of points with non-negative integer co-ordinates This particle starts at the origin 0 and at any point (x, y) travels to $(x + 1, y)$ if the event whose probability is p has occurred, otherwise to $(x, y + 1)$. Sampling terminates when the particle reaches a *boundary point*, and the set of such points is denoted by B. Any point which can be reached during sampling, including the boundary points, is *accessible*, and any path from the origin to a point B which can be traversed during sampling is *admissible*; all other points are *inaccessible* and all other paths *inadmissible*. The *index* of a point is the sum of its coordinates

It will probably help to note in particular that whereas Girshick, Mosteller and Savage used p to correspond to events causing the y co-ordinate to increase, we use it for x.

3. Determination of $N(x, y)$. The set B determines the sampling scheme and we are concerned with schemes in which all points of index greater than n, the finite maximum index of points in B, are inaccessible. This condition guarantees that if $N(x, y)$ denotes the number of admissible paths from the origin to a point (x, y) of B

$$\sum_B N(x, y)p^x(1 - p)^y \equiv 1,$$

the summation being over all boundary points. Consequently, to determine $N(x, y)$ equate coefficients of p in this identity, the coefficient of p^0 in the left hand side being 1 and all others zero. When all the $N(x, y)$ are known, the probability of reaching any subset of B can be calculated and the characteristics of the scheme found.

Sometimes it will be convenient to use

$$\sum_B N(x, y)q^y(1 - q)^x \equiv 1,$$

where $q = 1 - p$, but the resulting set of equations cannot be independent of the first set since if

$$\sum_{i=0}^{n} a_i p^i \equiv \sum_{j=0}^{n} b_j(1 - p)^j,$$

then

$$a_i = \sum_{j=i}^{n} (-1)^i \binom{j}{i} b_j.$$

The polynomial in either p or q is of degree n; the application of this method alone is therefore limited to boundaries containing at most $(n + 1)$ points, otherwise the number of unknowns exceeds the number of equations for them.

4. Properties of the boundary.

THEOREM 1. *If n is the maximum index of points in B and if any point of greater index is inaccessible, then B contains at least $n + 1$ points.*

There must be at least two boundary points of index n for any such point (a_n, b_n) must be approached from $(a_n - 1, b_n)$ or $(a_n, b_n - 1)$; in which case either $(a_n - 1, b_n + 1)$ or $(a_n + 1, b_n - 1)$ is a boundary point. Let P be any one of these points. At least one admissible path exists from 0 to P; suppose one such path to consist of the points (a_0, b_0), (a_1, b_1), $\cdots$, (a_n, b_n) where $a_k + b_k = k$ $(k = 0, 1, 2, \cdots, n)$. It is clear that one or more boundary points exist on the line $x = a_k$, having $y > b_k$, for otherwise the particle could travel indefinitely along this line; similarly one or more exist on $y = b_k$ with $x > a_k$, and if

there is just one on each they cannot be identical unless $k = n$ since (a_k, b_k) is not then a boundary point. Initially (a_0, b_0) contributes two boundary points; since then either $a_{k+1} = a_k$ and $b_{k+1} \neq b_k$ or $a_{k+1} \neq a_k$ and $b_{k+1} = b_k$ it follows that each succeeding point up to and including (a_{n-1}, b_{n-1}) contributes at least one more; the point (a_n, b_n) is counted as soon as x reaches a_n or y reaches b_n, whichever occurs first. Consequently there are at least $n + 1$ boundary points.

Reversely, if the boundary contains $n + 1$ points whose maximum index is m, such that any point of greater index is inaccessible, then $m \leq n$ For suppose $m > n$ and apply the preceding result.

An important class of boundaries therefore comprises those with the minimum number of points necessary to attain a given maximum index; they may conveniently be termed boundaries of minimum size and for them alone the method of equating coefficients yields the number of equations equal to the number of unknowns, the first being otherwise less than the second.

If there are exactly $n + 1$ boundary points then $(a_1, b_1), (a_2, b_2), \cdots, (a_{n-1}, b_{n-1})$ must each contribute to just one; since $a_{k+1} = a_k$ or $a_k + 1$ there is one point of B on each of the lines $x = 0, x = 1, \cdots, x = a_n$ and this set of points $(0, d_0)(1, d_1), \cdots, (a_n, b_n)$ can be denoted by U, the upper part of the boundary. Clearly $d_{k+1} \geq d_k - 1$ for otherwise more than one boundary point is required on the line $x = k + 1$. Similarly, there must be a second group of points of B $(c_0, 0), (c_1, 1), \cdots, (a_n, b_n)$ with $c_{k+1} \geq c_k - 1$ forming the lower boundary L; and all $(n + 1)$ points have now been enumerated, the point P belonging to both U and L. The characteristic of such sets B is that the sequences U and L both have monotonically non-decreasing index, the special case of sequences with monotonically increasing index provides the rejection and acceptance boundaries of non-rectifying industrial inspection procedures (The difference between rectifying and non-rectifying procedures is clearly stated in the introduction to Anscombe [1]).

THEOREM 2. *For boundaries of minimum size any two accessible points not in B of the same index m cannot be separated on the line $x + y = m$ by boundary or inaccessible points* In the terminology of Girshick, Mosteller and Savage the accessible points not in B form a *simple* region.

Let $Q(x_1, y_1)$ and $R(x_2, y_2)$ be any two such accessible points of index m and suppose $x_1 < x_2$. There are two possibilities: (a_m, b_m) does or does not lie between Q and R

(i) (a_m, b_m) lies between Q and R, i.e $x_1 < a_m < x_2$. In this case there must be points of B at $Q'(x_1, Y_1)$ with $Y_1 > y_1$ and at $R'(X_2, y_2)$ with $X_2 > x_2$ The boundary from Q' to P and from R' to P has non-decreasing index; hence all points of U on the lines $x = x_1, x = x_1 + 1, \cdots, x = a_m - 1$ have index at least $x_1 + Y_1 > m$; similarly all points of L on the lines $y = y_2, y = y_2 + 1, \cdots, y = b_m - 1$ have index at least $X_2 + y_2 > m$. By definition of the boundary there are no additional points of B on either group of lines between the path $0P$ and the line $x + y = n$, so the proof of the theorem is completed.

(ii) If $x_1 \geq a_m$ or $x_2 \leq a_m$ the proof is precisely analogous to that given in (i).

5. Justification of the method. THEOREM 3. *For boundaries of minimum size the equations for $N(x, y)$ are soluble and of rank $n + 1$.*

To prove this we give a general method of solution for the system of equations, using powers of p and q alternately: as already remarked, this is equivalent to using the equations from the coefficients of powers of p only. In the first place, note that the coefficient of q^u is a linear combination of numbers $N(x,y)$ with $x + y \geq u$ and $y \leq u$; and the coefficient of p^t has $x + y \geq t$ and $x \leq t$.

$$\text{Let } s = \text{Min}(d_0, d_1, d_2, \cdots, b_n) - 1.$$

Then from the coefficients of $q^0, q^1, \cdots, q^s$ can successively be determined $N(c_0, 0)\, N(c_1, 1), \cdots, N(c_s, s)$, the matrix of the equations being triangular with ones in the main diagonal. The points in U at $(r_1, s + 1), (r_2, s + 1), \cdots$ now appear in the coefficients of $q^{s+1}, q^{s+2}, \cdots$ and complicate the solution

$$\text{Let } r = \text{Max}(r_1, r_2, \cdots).$$

If either (r, d_r) or (c_s, s) is the point P then all the remaining $N(x, y)$ can successively be determined from the coefficients of powers of p when the values of $N(c_0, 0), N(c_1, 1), \cdots, N(c_s, s)$ are substituted in the equations. Otherwise the path OP for $y \geq s + 1$ must have $x \geq r + 1$ so that all points of L on $y \geq s + 1$ have $x \geq r + 2$ i.e. any point of L on $x = 0, x = 1, \cdots, x = r$ has $y \leq s$; for such points the number of admissible paths is now known. Therefore from the coefficients of $p^0, p^1, \cdots, p^r$ can successively be determined $N(0, d_0)$, $N(1, d_1), \cdots, N(r, d_r)$, the matrix of these unknowns being again triangular; in particular $N(r_1, s + 1), N(r_2, s + 1), \cdots$ can now be found.

Let $s_1 = \text{Min}\,(d_{r+1}, d_{r+2}, \cdots, b_n) - 1$, so that $s_1 > s$. The coefficients of $q^{s+1}, q^{s+2}, \cdots, q^{s_1}$ give successively $N(c_{s+1}, s + 1)\, N(c_{s+2}, s + 2), \cdots, N(c_{s_1}, s_1)$; for the points in U at $(r_{11}, s_1 + 1), (r_{12}, s_1 + 1) \cdots$. Let

$$r_1 = \text{Max}\,(r_{11}, r_{12}, \cdots).$$

Since there is only one point of U on each line $x = \text{constant}$, $r_1 > r$. As before, if either (r_1, d_{r_1}) or (c_{s_1}, s_1) is P the remaining points of U are soon determined Otherwise the process continues and there result an increasing sequence of points of L and a similar sequence for U; the process terminates when (a_n, b_n) has been reached in both, when all $N(x, y)$ will have been found

It is clear that for particular cases alternative methods of solution will prove more convenient.

6. Connection with estimation. Suppose that the point (t, u) is accessible and let $N^*(x, y)$ be the number of admissible paths from (t, u) to (x, y) where (x, y) is in B. Then Girshick, Mosteller and Savage have shown that $N^*(x, y)/N(x, y)$ is an unbiased estimate of $p^t(1 - p)^u$; and a necessary and sufficient condition for it to be the unique unbiased estimate is that the accessible points not in B form a simple finite region Hence from theorem 2 such estimates are unique for schemes with boundaries of minimum size. An alternative proof is given by

considering that if two unbiased estimates of any function of p exist and $f(x, y)$ is the difference between them at (x, y)

$$\sum_B f(x, y)N(x, y)p^x(1 - p)^y \equiv 0,$$

where $f(x, y)$ is not everywhere zero. The equations formed by equating coefficients have rank $(n + 1)$ as shown by Theorem 3, so that the only solution is $f(x, y)N(x, y) = 0$ Since each $N(x, y)$ is certainly positive it follows at once that $f(x, y) = 0$ and there can only be one unbiased estimate.

7. An illustration. As an application of the method we take the interesting rectifying sequential inspection scheme discussed by Anscombe. The boundary points are at $(H, 0)$, $(H + b, 1)$, $\cdots$ $(H + \mu b, \mu)$, where μ is the greatest integer less than $(N - H)/(b + 1)$, and thereafter on the line $x + y = N$. The equations for $N(x, y)$ take here their simplest form, namely equation (4) of Barnard's paper. From the coefficients of $q^0, q^1, \cdots, q^y, \cdots,$

$$1 = N(H, 0);$$

$$0 = N(H + b, 1) - HN(H, 0) \quad \text{whence} \quad N(H + b, 1) = H;$$

$$0 = N(H + 2b, 2) - \binom{H + b}{1} H + \binom{H}{2} \quad \text{whence} \quad N(H + 2b, 2) = \frac{H(H + 2b + 1)}{2!},$$

$$0 = N(H + 3b, 3) - \binom{H + 2b}{1} \frac{H(H + 2b + 1)}{2!} - \binom{H + b}{2} H + \binom{H}{3};$$

$$\text{whence } N(H + 3b, 3) = \frac{H(H + 3b + 2)(H + 3b + 1)}{3!}.$$

It now appears reasonable to guess the general term as

$$\frac{H}{y!}(H + yb + y - 1)(H + yb + y - 2) \cdots (H + yb + 1).$$

The proof is therefore complete if we show

$$\binom{H}{y} - \binom{H + b}{y - 1} H + \binom{H + 2b}{y - 2} \frac{H(H + 2b + 1)}{2!} - \binom{H + 3b}{y - 3} \frac{H(H + 3b + 2)(H + 3b + 1)}{3!}$$
$$+ \cdots + (-1)^y \frac{H(H + yb + y - 1)(H + yb + y - 2) \cdots (H + yb + 1)}{y!} = 0.$$

Put $(b + 1) = \xi$, and the left hand side becomes

$$\frac{(H-1)!}{(H-y)!y!} - \frac{(H+\xi-1)!}{(H+\xi-y)!(y-1)!1!} + \frac{(H+2\xi-1)!}{(H+2\xi-y)!(y-2)!2!}$$
$$- \cdots (-1)^y \frac{(H+y\xi-1)!}{(H+y\xi-y)!y!},$$

which is y times the coefficient of t^{H-y} in $(1+t)^{H+y\xi-1} \times [(1+t)^{-\xi} - t^{-\xi}]^y$. Rewriting the latter as $(1+t)^{H-1}[1-(1+t^{-1})^{\xi}]^y$, it becomes clear that the highest power of t is t^{H-y-1}, whence the required result follows.

REFERENCES

[1] F. J. Anscombe, "Linear sequential rectifying inspection for controlling fraction defective," *Roy. Stat. Soc. Jour. (supplement)*, Vol. 8 (1946), pp 216–222.

[2] G A. Barnard, "Sequential tests in industrial statistics," *Roy Stat. Soc Jour (supplement)*, Vol. 8 (1946), pp 1–21.

[3] J. P. Burman, "Sequential sampling formulas for a binomial population," *Roy Stat. Soc Jour. (supplement)*, Vol. 8 (1946), pp. 98–103.

[4] M. A Girshick, F. Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems with applications," *Annals of Math Stat*, Vol 17 (1946), pp. 13–23.

[5] J. B. S. Haldane, "On a method of estimating frequencies," *Biometrika*, Vol 33 (1945), pp. 222–225.

[6] C. M. Stockman and P. Armitage, "Some properties of closed sequential schemes," *Roy. Stat. Soc. Jour. (supplement)*, Vol. 8 (1946), pp. 104–112.

[7] J. Wolfowitz, "On sequential binomial estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 489–492.

This section is devoted to brief research and expository articles and other short items

NON-PARAMETRIC TOLERANCE LIMITS[1]

By R. B. Murphy

Princeton University

1. Summary. In this note are presented graphs of minimum probable population coverage by sample blocks determined by the order statistics of a sample from a population with a continuous but unknown cumulative distribution function (c.d.f.). The graphs are constructed for the three tolerance levels .90, .95, and .99. The number, m, of blocks excluded from the tolerance region runs as follows: $m = 1(1)6(2)10(5)30(10)60(20)100$, and the sample size, n, runs from m to 500.

Thus the curves show the solution, β, of the equation $1 - \alpha = I_\beta(n - m + 1, m)$ for $\alpha = .90, .95, .99$ over the range of n and m given above, where $I_x(p, q)$ is Pearson's notation for the incomplete beta function

Examples are cited below for the one- and two-variate cases. Finally, the exact and approximate formulae used in computations for these graphs are given

2. Introduction. Suppose a sample of size n is drawn from a population having a continuous cumulative distribution function (c d f), $F(x)$. Let the sample values arranged in order of increasing magnitude be $x_1, x_2, \cdots, x_n$. The fraction, u, of the population which is included between x_r (the r-th smallest value in the sample) and x_{n-s+1} (the s-th largest value) is $F(x_{n-s+1}) - F(x_r)$ This quantity u has been called the *population coverage* for the interval (x_r, x_{n-s+1}) The probability element for this coverage is

$$f(u)\,du = \frac{\Gamma(n+1)}{\Gamma(n-m+1)\Gamma(m)} u^{n-m}(1-u)^{m-1}\,du \tag{2.1}$$

where $m = r + s$. From (2.1) we can calculate the probability that this coverage is at least a given amount, say β. If we call this probability α, we have

$$\alpha = \int_\beta^1 f(u)\,du. \tag{2.2}$$

The quantity α is the probability that $100\beta\%$ of the population will be included between x_r and x_{n-s+1}, and it is called the *tolerance level*. This probability depends only on n and m $(=r + s)$.

---

[1] All computations involved in this paper were carried out under an Office of Naval Research contract.

The idea of coverage is more general than it first appears. If we think of $x_1, x_2, \cdots, x_n$ as points plotted along the x-axis, we will then have $n + 1$ intervals: $(-\infty, x_1), (x_1, x_2), \cdots, (x_n, +\infty)$, which, following Tukey [3], we will call *blocks*. The reason for this term will be clear when we deal with the case of a sample from a population of more than one variable. The coverage for the i-th block (x_i, x_{i+1}) is $F(x_{i+1}) - F(x_i)$. The probability element of the sum of the coverages of *any* preassigned group of $n - m + 1$ blocks is given by (2.1) and hence the probability α that the fraction of the population covered by *any* $n - m + 1$ blocks is given by (2.2). By preassigned blocks we mean ones designated by order statistics prior to obtaining any sample from which a prediction is to be made with these blocks In general it is *not* legitimate, after taking a sample and for some reason evident only then, to specify which blocks in this sample are to be included or excluded from the coverage. There is no objection, however, to specifying a scheme of blocks for the coverage on the basis of past samples when the scheme is to be applied to future samples.

The purpose of this note is to present graphs of β as a function of n for $m = 1(1)6(2)10(5)30(10)60(20)100$ and for $\alpha = .90, .95, .99$ There are three figures: Figure 1 gives curves for $\alpha = .90$, Figure 2 for $\alpha = 95$, and Figure 3 for $\alpha = .99$. The graphs are accurate to at least two decimal places but never more than three. In terms of the Pearson notation (2.2) gives, after minor alternation, $1 - \alpha = I_\beta\ (n - m + 1, m)$. Hence these graphs may also be used to find the 10, 5 and 1 per cent points of a variate X $(0 \leq X \leq 1)$ with the c.d.f. $I_x(p, q)$ for $1 \leq p \leq 500$ and $1 \leq q \leq 100$.

3. Computations for the graphs. If in the relation (2.2) three of the arguments α, β, m, and n are given, the solution for the fourth may often be found in Pearson [5] or Thompson [6]. The values of β through $n = 100$ were computed exactly for these graphs. For larger n, β was computed approximately from

$$\beta \cong \left[\frac{\sqrt{(\chi_\alpha^2 - 2m)^2 + 16n(n - m)} - (\chi_\alpha^2 - 2m)}{4n}\right]^2 \tag{3.1}$$

where χ_α^2 is determined by the relation

$$Pr(\chi^2 \geqq \chi_\alpha^2) = 1 - \alpha$$

and has $2m$ degrees of freedom. This approximation is due to Scheffé and Tukey. For large m the Cornish-Fisher approximation to χ_α^2 was used.

4. Illustrations of the one-variate case. The most common use to which the graphs presented here may be put is in the prediction of β in sampling from a distribution of a single random variable. It is this case that was first presented by Wilks [1]. Suppose in the mass production of a certain type of screw one is interested in the least proportion of all screws manufactured that have lengths between the least and greatest lengths appearing in a random sample of 100

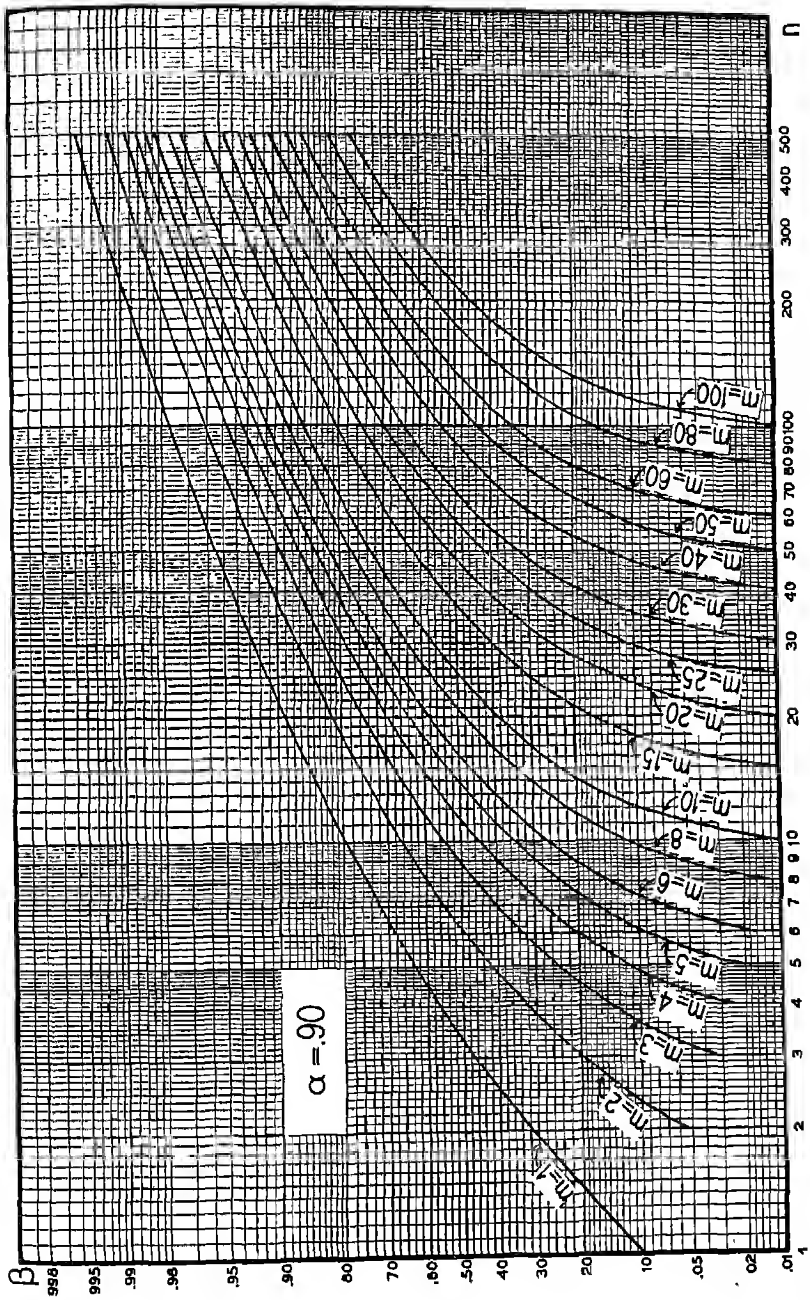


FIG 1 Graphs of Population Coverage for the Tolerance Level .90

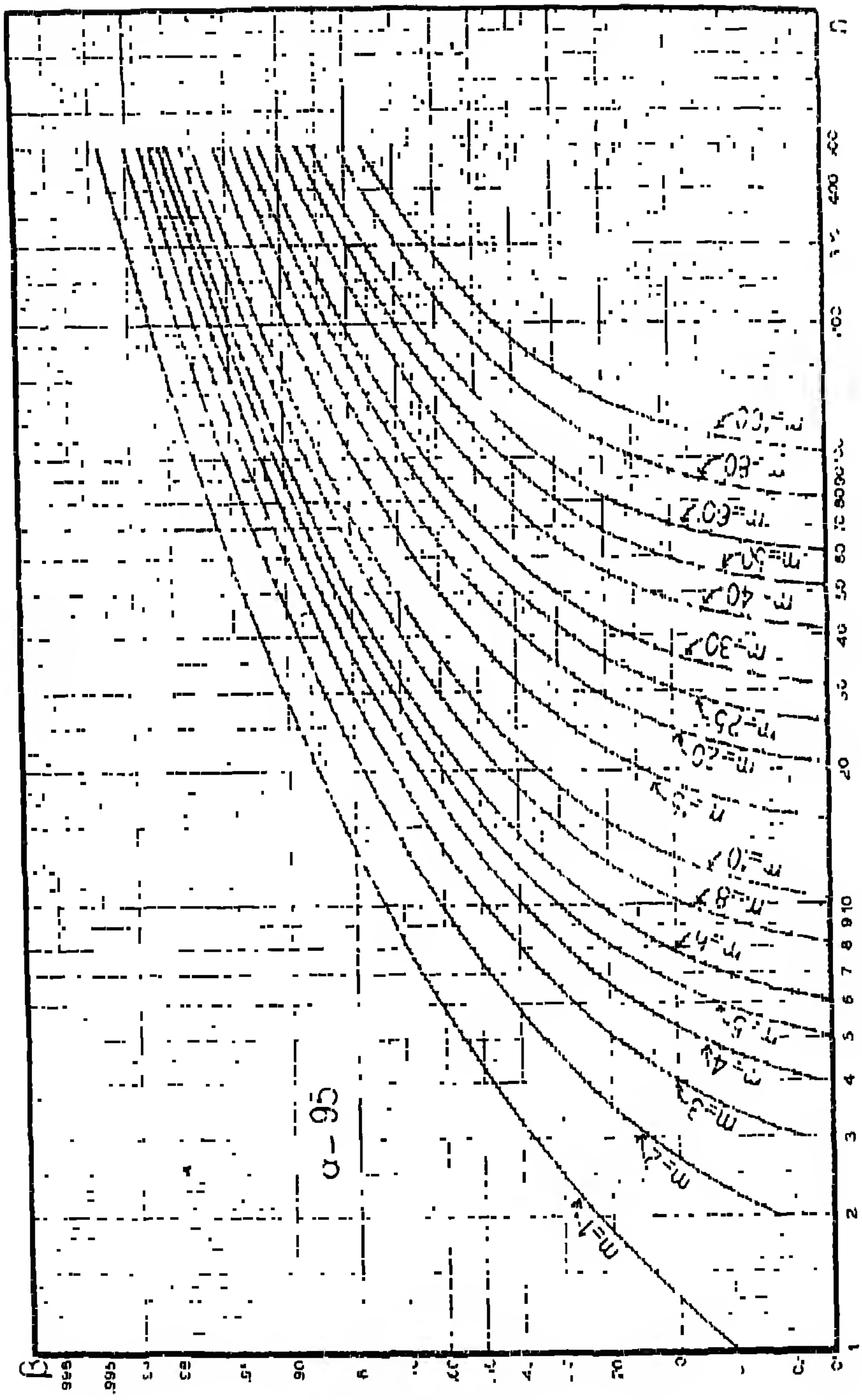


FIG. 2 Graphs of Population Coverage for the Tolerance Level .95.

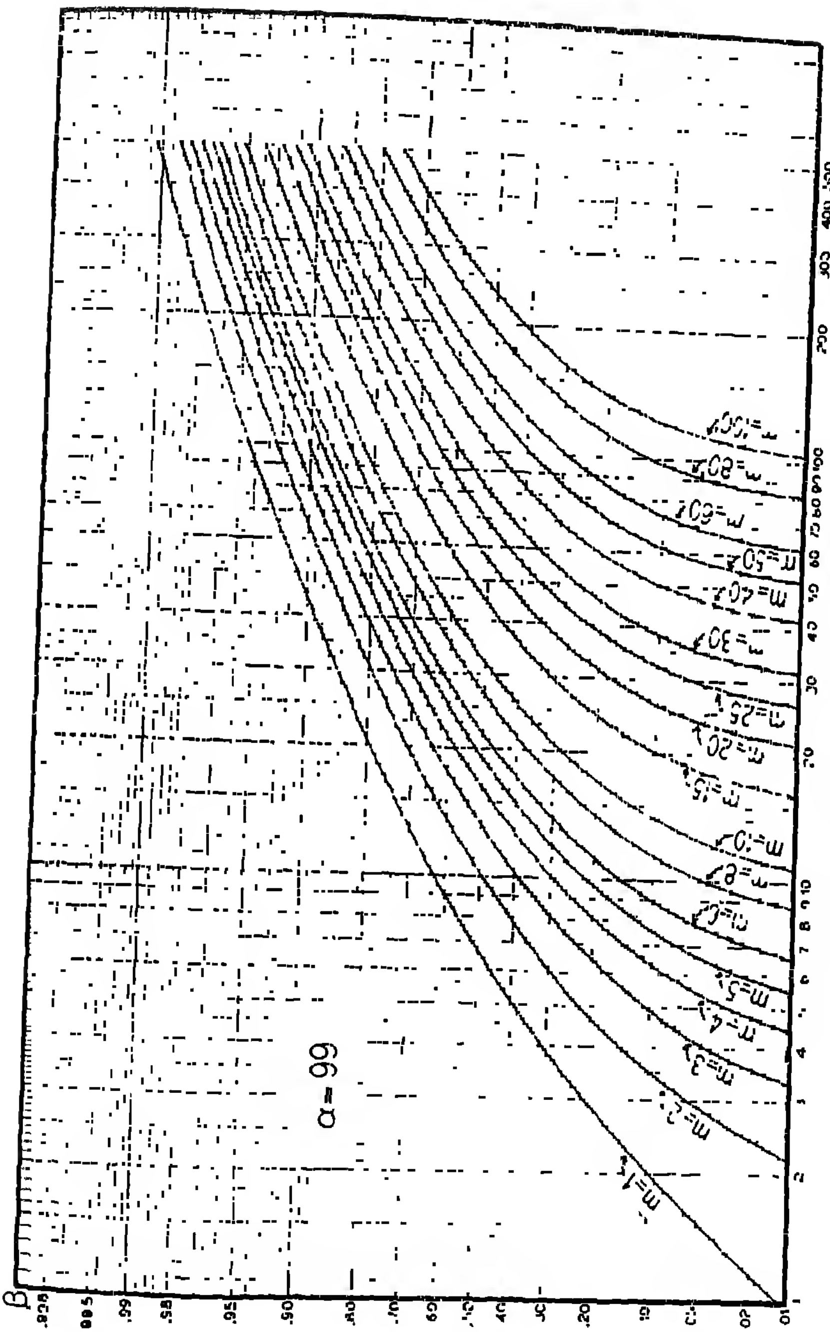


Fig 3 Graphs of Population Coverage for the Tolerance Level 99

screws. It is assumed that we do not know the distribution of the length, X, of a screw produced in this process. Furthermore, it is assumed, of course, that the manufacturing process is in a state of statistical control in the sense of Shewhart. We plan to discard two blocks: $(-\infty, x_1)$ and $(x_{100}, +\infty)$—exactly as many blocks as observations. At the level $\alpha = .99$ we obtain from Figure 3 that at least 93.5% of all screws in the population sampled have lengths that fall between x_1 and x_{100}. If we now draw a random sample of 100 screws and find the least and greatest screw lengths to be 1.40 and 1.60 inches respectively, we may say that at least 93.5% of all screws from the population sampled have lengths between 1.40 and 1.60 inches at the .99 tolerance level. It must be observed that the prediction is made on the basis of preassigned order statistics, and not of the values 1.40 and 1.60.

We might equally as well have put the question in another way: If we want at least 93.5% of the lengths of all screws to lie within the range of lengths of a sample of 100 screws, then at the tolerance level $\alpha = .99$ what is the smallest sample we could have in which as many as 2% of the sample are not acceptable? Examining the intersections of the curves in Figure 3 with the line $\beta = .935$ we choose the smallest n such that $m/n \leqq .02$ and find $n = 100$.

5. The case of more than one variate. The ideas given in the introduction may be extended to sampling situations involving two or more statistically dependent variates with a continuous joint c.d.f. by means of the notion of blocks. The abstract formulation is given by Tukey [3]. We shall restrict ourselves to the case of two dependent variates X and Y, but the generalization is obvious. Because of the dependence, the joint population of X and Y may be expressed as an associated pair of values $W = (X, Y)$. Suppose a sample of size n is drawn from this population, and let the pairs be $w_1, w_2, \cdots, w_n$, where $w_i = (x_i, y_i)$. If we now choose a sequence of n numerically valued functions of x and y (or of w), $f_1(w), \cdots, f_n(w)$, let us order the w_i in a sequence $w_1^{(1)}, w_2^{(1)}, \cdots, w_n^{(1)}$ such that $f_1(w_{i+1}^{(1)}) > f_1(w_i^{(1)})$. Imagine now that the sample values are plotted in a plane scatter diagram. We call the first block the set of points $w = (x, y)$ such that $f_1(w) < f_1(w_1^{(1)})$. That is, we may imagine the curve $f_1(x, y) - f_1(w_1^{(1)}) = 0$ plotted in the plane and that the first block is bounded by this curve. Then discarding $w_1^{(1)}$ we take the $n - 1$ remaining w_i and order them in a sequence $w_1^{(2)}, w_2^{(2)}, \cdots, w_{n-1}^{(2)}$ such that $f_2(w_{i+1}^{(2)}) > f_2(w_i^{(2)})$. We call the second block the set of points $w = (x, y)$ such that $f_1(w) \geq f_1(w_1^{(1)})$ and also $f_2(w) < f_2(w_1^{(2)})$. Thus the second block is bounded by the curves $f_1(x, y) - f_1(w_1^{(1)}) = 0$ and $f_2(x, y) - f_2(w_1^{(2)}) = 0$. If we continue this process of discarding and reordering, until all n functions f_i are used, we shall obtain a division of the plane into $n + 1$ non-overlapping blocks, the "extra" block arising at the last step in the process. Then the fraction, u, of "points" (X, Y) of the joint population of X and Y that are covered by *any* $n - m + 1$ blocks has the probability element (2.1). Also the probability α that the population coverage, u, will be at least as large as β is given by (2.2). The $n - m + 1$ blocks constitute a tolerance *region*.

An extension of this case has been made by Wald [2]. Namely, before a sample is taken let us choose a numerically valued function f of w and choose $k(\leq n)$ of the w_i and order them in a sequence $w_{a_1}^{(0)}, w_{a_2}^{(0)}, \cdots, w_{a_k}^{(0)}$ such that $f(w_{a_{j+1}}^{(0)}) > f(w_{a_j}^{(0)})$ and $a_{j+1} > a_j$. Next, within each "strip" of the (x, y) plane such that $w = (x, y)$ satisfies $f(w_{a_{j+1}}^{(0)}) > f(w) > f(w_{a_j}^{(0)})$, suppose that we follow the construction in the previous paragraph. Then the population coverage, u, by $n - m + 1$ blocks from one or more of these strips or their exteriors has the probability element (2.1).

Again the warning must be made that the above functions $f, f_1, f_2, \cdots, f_n$, the numbers $a_1, a_2, \cdots, a_k$ and the sequence of construction must be completely specified before samples are drawn to which this scheme is to be applied.

6. Illustrations for two variates. As an example of the use of the graphs for a two-variate case, we use an example cited by Tippett [8]. The two variates are the percentage of pig iron, X, and the lime consumption, Y, per cwt. of steel in 100 steel castings made without slag control. A scatter diagram is given in Figure 4. Unfortunately the value of this example is lessened by the fact that the block schemes were made after the sample had been taken; it does illustrate, at least, the two simple types of scheme.

The tolerance region T (solid lines in Figure 4) resulted from the following scheme: let $f_1(w) = y, f_2(w) = f_3(w) = f_4(w) = f_5(w) = f_6(w) = -y$. Now follow the Wald procedure choosing $f(w) = y$ with $k = 6$, and $a_1 = 1, a_2 = 13, a_3 = 46, a_4 = 75, a_5 = 90, a_6 = 96$. Then in each strip $y_{a_{j+1}} > y > y_{a_j}$ let $f_i(w) = x$. Considering only the blocks within the heavy line as the tolerance region, we have, by counting the discarded blocks, $m = 16$.

In constructing the region T' (broken lines in Figure 4) we also use Wald's method, taking $f(w) = y - 5x$ with $k = 2$ and $a_1 = 3, a_2 = 96$ In the exterior region with $f(w) > f(w_{96}^{(0)})$ let all $f_i = y + 5x$ and similarly in the exterior region $f(w) < f(w_3^{(0)})$. Then in the strip $f(w_{96}^{(0)}) > f(w) > f(w_3^{(0)})$ (i.e., in the region in which $41 > y - 5x > -77$) choose $f_1(w) = y, f_2(w) = f_3(w) = f_4(w) = -y, f_5(w) = f_6(w) = f_7(w) = y + 5x$, and $f_8(w) = f_9(w) = -y - 5x$. Counting the blocks outside the heavily bordered region, we have $m = 17$.

We obtain by interpolation $\beta = .80$ for T and $\beta = .78$ for T' at the $\alpha = .90$ level

7. Ties. A tie is a sample point which in a coordinate system defining a set of order statistics coincides in one or more coordinates with other sample points. For instance, in the X coordinate of our example (32, 159) and (32, 185) are tied, and (47, 218) and (47, 218) are tied in any system of coordinates. It would seem easier to avoid ties with regions of the type of T' than with those of the type of T.

The existence of ties in the population is assumed impossible, because positive point probabilities would destroy the continuity of the c.d.f Therefore we attribute the ties to the crudity of measuring devices.

A procedure for handling ties is given by Tukey [4].

8. Acknowledgments. The author wishes gratefully to acknowledge the assistance of Dr. S. S. Wilks in the preparation of this note and of Dr. J. W. Tukey, who also suggested the data used in section 6.

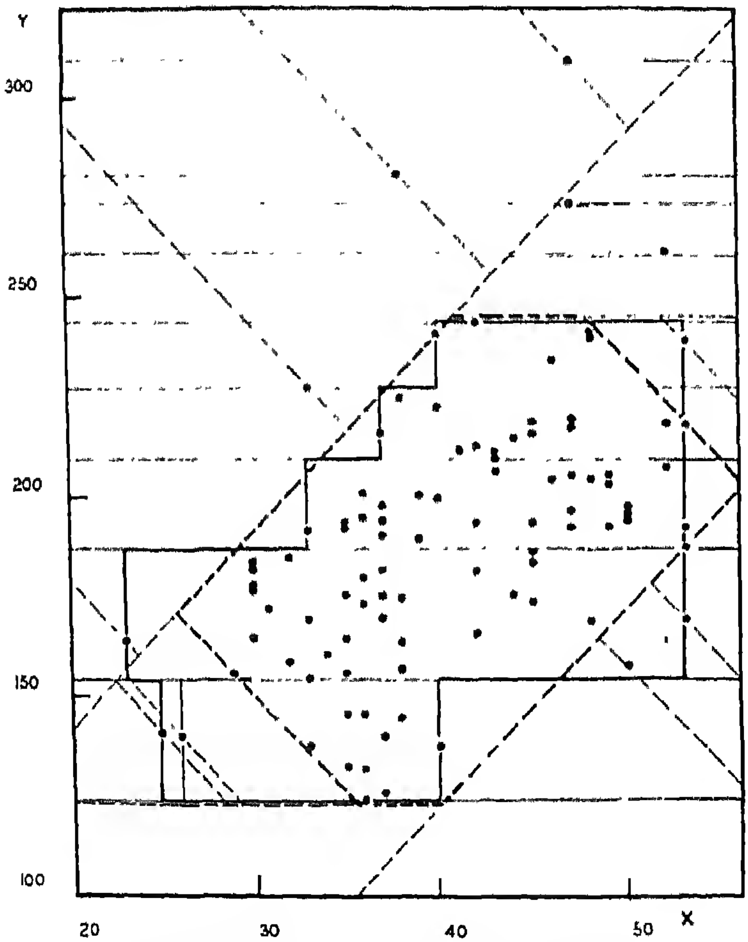


FIG. 4. Illustrative Tolerance Regions for Two Variates.

REFERENCES

[1] S. S. WILKS, "Statistical prediction with special reference to the problem of tolerance limits," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 400–409

[2] A. WALD, "An extension of Wilks' method for setting tolerance limits," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 45–55.

[3] J. W. TUKEY, "Non-parametric estimation II. Statistically equivalent blocks and tolerance regions—the continuous case," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 529–539.

[4] J. W. TUKEY, "Non-parametric estimation III. Statistically equivalent blocks and multivariate tolerance regions the discontinuous case," *Annals of Math Stat.*, Vol. 19 (1948), pp. 30–39.

[5] K. PEARSON, *Tables of the Incomplete Beta-Function*, Cambridge, 1934.

[6] C. M. THOMPSON, "Tables of percentage points of the incomplete beta function," *Biometrika*, Vol. 32, Part II (1941), pp. 151–181.

[7] H. GOLDBERG AND H. LEVINE, "Approximation formulas for the percentage points and normalization of t and χ^2", *Annals of Math. Stat.*, Vol 17 (1946), pp. 216–225

[8] L. H. C. TIPPETT, *Statistical Methods in Industry*, Iron and Steel Industrial Research Council, British Iron and Steel Federation, 1943

THE FOURTH DEGREE EXPONENTIAL DISTRIBUTION FUNCTION[1]

BY LEO A. AROIAN

Hunter College

We shall derive a recursion formula for the moments of the fourth degree exponential distribution function, state its more characteristic features, and show how the graduation of observed distributions may be accomplished by the method of moments and the method of maximum likelihood. The purpose of the note is to make possible a wider use of this function.

R. A. Fisher [1] introduced the fourth degree exponential function

$$y_t = k \exp\{-(\beta_4 t^4 + \beta_3 t^3 + \beta_2 t^2 + \beta_1 t)\}, \tag{1}$$

where $r_1 \leq t \leq r_2$, $t = (x - m)/\sigma$, m indicates the population mean, σ the population standard deviation, and where the β's are functions of

$$\alpha_n = \int_{r_1}^{r_2} t^n y_t \, dt.$$

A. L. O'Toole in two stimulating papers [2], [3], has studied (1), however his methods and results are unnecessarily complicated. O'Toole requires eight moments to determine parameters similar to the β's. Both Fisher and O'Toole considered the restricted class of (1) with range $(-\infty, \infty)$

Let

$$u = t^n \exp\{-(\beta_4 t^4 + \beta_3 t^3 + \beta_2 t^2)\}, \; dv = e^{-\beta_1 t}\, dt \tag{2}$$

in

$$\alpha_n = \int_{r_1}^{r_2} t^n y_t \, dt, \quad \text{obtaining} \tag{3}$$

$$4\beta_4\alpha_{n+3} + 3\beta_3\alpha_{n+2} + 2\beta_2\alpha_{n+1} + \beta_1\alpha_n = n\alpha_{n-1}, \qquad n = 1, 2, 3, \cdots, \tag{4}$$

[1] Presented to the American Mathematical Society and the Institute of Mathematical Statistics, September 4, 1947.

and for $n = 0$, the right side of (4) is defined as zero. The result (4) is valid under the assumption

$$uv\Big]_{r_1}^{r_2} = 0. \tag{5}$$

Given the first six moments, β_4, β_3, β_2, β_1 are readily determined. It will be found that if $\beta_4 > 0$, $\beta_3 \neq 0$, then $r_1 = -\infty$, $r_2 = \infty$; while if $\beta_4 < 0$, and $\beta_3 \neq 0$, r_1 and r_2 will be finite. If we set $n = 0, 1, 2, 3$, in (4), the solutions are

$$
\begin{aligned}
\beta_4 &= \{\alpha_3(\alpha_5 - 4\alpha_3) - (\alpha_4 - 3)(\alpha_4 - 1)\} \div 4D; \\
\beta_3 &= \{-\alpha_3(\alpha_6 - 3\alpha_4 - \alpha_3^2) + (\alpha_5 - \alpha_3)(\alpha_4 - 3)\} \div 3D; \\
\beta_2 &= \{(\alpha_3 - \alpha_5)(\alpha_5 - 4\alpha_3) + (\alpha_4 - 1)(\alpha_6 - \alpha_3^2 - 3\alpha_4)\} \div 2D; \\
\beta_1 &= \{\alpha_3(\alpha_6 - \alpha_3\alpha_5 - 3\alpha_4 + 3\alpha_3^2) - (\alpha_4 - 3)(\alpha_5 - \alpha_3\alpha_4)\} \div D,
\end{aligned} \tag{6}
$$

where

$$D = (\alpha_6 - \alpha_4^2 - \alpha_3^2)(\alpha_4 - \alpha_3^2 - 1) - (\alpha_5 - \alpha_3 - \alpha_3\alpha_4)^2 \geqq 0.$$

To prove $D \geqq 0$ we adopt the method of J. E. Wilkins Jr. [4]. In only a trivial case is $D = 0$. Let

$$G(a, b, c, d) = \int_{r_1}^{r_2} (a + bt + ct^2 + dt^3)^2 y_t \, dt \geqq 0,$$

where y_t is any probability function with range $r_1 \leqq t \leqq r_2$. Since $G(a, b, c, d)$ is a semi-definite quadratic form, its discriminant will be non-negative. But its discriminant is easily seen to be equal to D, thus

$$D = \begin{vmatrix} \alpha_3 & 1 & 0 & 1 \\ \alpha_4 & \alpha_3 & 1 & 0 \\ \alpha_5 & \alpha_4 & \alpha_3 & 1 \\ \alpha_6 & \alpha_5 & \alpha_4 & \alpha_3 \end{vmatrix} \geqq 0. \tag{7}$$

We summarize without proofs the essential features of the fourth degree exponential. Near the normal point, $\alpha_4 = 3$, $\alpha_3 = 0$, the fourth degree exponential function, the Pearson system, and the Gram-Charlier Type A are essentially alike. Type C [5] while similar is not the same. Note that β_4 may be negative and in such a case r_1 and r_2 are the two real zeros of the derivative of (1). The exponential may be bimodal as well as unimodal and the normal curve is the special case $\beta_4 = \beta_3 = \beta_1 = 0$. Various special cases where a particular β is zero are readily handled by either (4) or (6). The graduation of both unimodal and bimodal observed distributions will be published elsewhere.

Let

$$y_t = k \exp - \sum_{j=1}^{r} \beta_j t^j, \qquad r_1 \leq t \leq r_2, \tag{8}$$

where

$$\frac{1}{k} = \int_{r_1}^{r_2} \exp - \sum_{j=1}^{r} \beta_j t^j \, dt. \tag{9}$$

The likelihood, L, in a sample of N is given by

$$L = k^N \exp\left\{-\left\{\beta_r \sum_{i=1}^{N} t_i^r + \beta_{r-1} \sum_{i=1}^{N} t_i^{r-1} + \cdots + \beta_1 \sum_{i=1}^{N} t_i\right\}\right\} \tag{10}$$

where $t_i = (x_i - m)/\sigma$. Then

$$\frac{\partial \log L}{L\, \partial\beta_j} = \frac{N}{k}\frac{\partial k}{\partial\beta_j} - \sum t_i^j, \quad \text{and} \tag{11}$$

$$\begin{aligned}\frac{1}{k}\frac{\partial k}{\partial\beta_j} = k\left\{\int_{r_1}^{r_2} t^j \exp\left\{-\sum_{j=1}^{r} \beta_j t^j\, dt\right\} - \frac{\partial r_2}{\partial\beta_j} \exp\left\{-\sum_{j=1}^{r} \beta_j r_2^j\right\}\right. \\ \left. + \frac{\partial r_1}{\partial\beta_j} \exp\left\{-\sum_{j=1}^{r} \beta_j r_1^j\right\}\right\}.\end{aligned} \tag{12}$$

If we assume either r_1 and r_2 constant, or $\exp\left\{-\sum_{j=1}^{r} \beta_j r_2^j\right\}$ and $\exp\left\{-\sum_{j=1}^{r} \beta_j r_1^j\right\}$ negligible, then (12) becomes

$$k \int_{r_1}^{r_2} t^j \exp\left\{-\sum_{j=1}^{r} \beta_j t^j\right\} dt \quad \text{and} \quad \frac{\partial \log L}{L\,\partial\beta_j} = 0 \quad \text{implies}$$

$$\frac{\int_{r_1}^{r_2} t^j \exp\left\{-\sum_{j=1}^{r} \beta_j t^j\right\} dt}{\int_{r_1}^{r_2} \exp\left\{-\sum_{j=1}^{r} \beta_j t^j\right\} dt} = \frac{\sum t_i^j}{N} = a_j, \qquad j = 1, 2, \quad r, \tag{13}$$

where a_j is the sample estimate of α_j. For, if in $\sum t_i^j/N$ we let $j = 1, 2$, we find by (13) that $\bar{x} = m$, and $\sigma^2 = \sum(x_i - \bar{x})^2/N$ The solution of (13) provides estimates of β_4, β_3, β_2, and β_1, if we set $r = 4$. Naturally more time is required for the solution of (13) as compared with the method of moments, but the maximum likelihood estimates are asymptotically efficient. The system (13) must be solved by successive approximations. To determine the moments solution all we do is to replace α_j by a_j in equations (6). This affords a point of departure from which the maximum likelihood equations may be solved. The two methods are not the same.

The fourth degree exponential is readily generalized to a fourth (or rth) degree multivariate function including the normal multivariate function as a special case.

REFERENCES

[1] R. A. Fisher, "The mathematical foundations of theoretical statistics." *Roy Soc. Phil Trans.*, Vol. 222, Series A, (1922), pp 355–56 particularly

[2] A. L. O'TOOLE, "On the system of curves for which the method of moments is the best method of fitting," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 1–29.
[3] A. L. O'Toole, "A method of determining the constants in the bimodal fourth degree exponential function," *Annals of Math. Stat.*, Vol. 4 (1933), pp. 79–93.
[4] J. ERNEST WILKINS, JR., "A note on skewness and kurtosis," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 333–335.
[5] C. V. L. CHARLIER, "A new form of the frequency function," *Lund universitet, Acta*, N. F. Bd 24 (1928), Avd. 2, Art. no. 8, pp. 1–28.
[6] M. G. KENDALL, *The Advanced Theory of Statistics*, Griffin & Co., London, Vol. II, p. 43.

AN APPROXIMATION TO THE BINOMIAL SUMMATION

BY G. F. CRAMER

Washington, D. C.

We consider the binomial expansion $(q + p)^n$, where $q = 1 - p$ and n is a positive integer. For given values of n, p, r, and s, where $np < r < s \leq n$, we are often interested in the probability $P(r \leq x \leq s)$ that the number of successes x will satisfy $r \leq x \leq s$.

When n does not exceed 50, we can use tables of the Incomplete Beta Function, or other convenient and accurate tables. For "large" values of n, we can use normal tables. When p is "small", we can use Poisson tables. However, it is often true that p is fairly small, and yet not small enough to give really accurate results when Poisson tables are employed in the usual way, while n is too large for use of the tables of the Incomplete Beta Function and yet too small for accurate use of normal tables.

It frequently happens that an upper bound for $P(r \leq x \leq s)$ would serve our purpose. We propose to show how to find this from Poisson tables with greater accuracy than could be obtained by using these tables in the ordinary way.

We shall denote the general term of the binomial expansion by $B_i = \binom{n}{i}p^iq^{n-i}$ and the general term of the corresponding Poisson distribution with the same value of p by $P_i = (pn)^ie^{-pn}/i!$. We shall also consider a second Poisson distribution whose general term is given by $P_i' = (p'n)^ie^{-p'n}/i!$, where $p' \neq p$ will be determined later.

We shall use the following notations:

$$U_i = B_{i+1}/B_i = (n - i)(p)/(i + 1)(1 - p); \tag{1}$$

$$V_i = P_{i+1}/P_i = pn/(i + 1); \tag{2}$$

$$V_i' = P_{i+1}'/P_i' = p'n/(i + 1); \tag{3}$$

$$U_i - V_i = p(np - i)/(i + 1)(1 - p). \tag{4}$$

From (4) we obtain at once the following:

LEMMA I. *$U_i > V_i$ or $U_i < V_i$ according as $i < np$ or $i > np$.*

Thus, the size of the general term of the binomial expansion falls off more steeply to the right of $i = np$ than does that of the general Poisson term.

We can use lemma I to obtain an upper bound to $P(r \leq x \leq s)$ for any $r > np$. In fact,

$$
\begin{aligned}
&B_r = B_r P_r / P_r ;\\
&B_{r+1} < B_r P_{r+1} / P_r ;\\
&B_{r+2} < B_{r+1} P_{r+2} / P_{r+1} < B_r P_{r+2} / P_r ;\\
&\quad\vdots\\
&B_s < B_r P_s / P_r .
\end{aligned}
$$

Adding these, we obtain

$$P(r \leq x \leq s) = \sum_{i=r}^{s} B_i < (B_r/P_r) \sum_{i=r}^{s} P_i = (B_r/P_r)\left(\sum_{i=r}^{\infty} P_i - \sum_{i=s}^{\infty} P_i\right) \tag{5}$$

The quantity in parentheses in (5) can be found by use of the cumulative Poisson table provided, of course, it is within the range of that table, while the B_r/P_r can be computed directly.

In the work we have done so far, we have used a Poisson distribution which is less steep than the corresponding binomial distribution throughout the whole interval $np < r \leq x \leq n$. It seems reasonable to investigate the possibility of improving upon (5) by using a Poisson distribution having a different value p' in place of p, where p' is chosen so that the new Poisson distribution is of the same steepness at $x = r$ as is the binomial distribution. We wish to have $U_r = V'_r$ and $U_i \leq V'_i$ for all $r \leq i \leq n$. The first of these conditions requires that $(n - r)(p)/(r + 1)(1 - p) = p'n/(r + 1)$. Solving for p' we obtain

$$p' = (n - r)(p)/(n)(1 - p). \tag{6}$$

We are now ready to prove the following:

LEMMA II. *If p' is defined by* (6) *and if U_i, V_i, and V'_i are defined by* (1), (2), *and* (3) *respectively, then $U_i \leq V'_i < V_i$, provided $r > np$ and $i \geq r$.*

It is easy to see that $U_i/V'_i = (n - i)(p)(1 + i)/(1 + i)(1 - p)(np')$, and this can be reduced to $(n - i)/(n - r)$ by replacing p' by its value from (6). Then $U_i/V'_i \leq 1$ since $i \geq r$. Moreover, we have $V'_i/V_i = (p'n)(i + 1)/(i + 1)(pn) = p'/p = (n - r)/(n - np)$. But $r > np$ and hence $V'_i < V_i$. This completes the proof of Lemma II.

We are now in a position to obtain an inequality somewhat better than (5). The derivation of the new upper bound for $P(r \leq x \leq s)$ goes just as before except that each P_i is replaced by P'_i. We obtain the new inequality

$$P(r \leq x \leq s) < K'B_r/P'_r, \tag{7}$$

where $K' = \sum_{i=r}^{\infty} P'_i - \sum_{i=s}^{\infty} P'_i$.

We can get a lower bound as well as a somewhat improved upper bound for

$P(r \leq x \leq s)$ by calculating B_r and B_{r+1} directly and then applying (5) or (7) to find an upper bound M of $P(r+1 \leq x \leq s)$. This gives the inequality

$$(8) \qquad B_r + B_{r+1} < P(r \leq x \leq s) < B_r + M.$$

This could, of course, be still further improved by calculating directly still more of the B_i's and using a similar procedure, but one would not care to carry this very far.

To illustrate the various approximations, we have worked out a numerical example the results of which appear below. For convenience in checking, we have used a value of n which is within the range of the tables of the Incomplete Beta Function, even though we would ordinarily use our method only for larger values of n.

Example. $s = n = 40$; $r = 10$; $p = 1/10$; $p' = 1/12$. The tables of the Incomplete Beta Function give $P(10 \leq x \leq 40) = .0050631$. Using Poisson tables in the usual way, we get $P(10, 4) - P(40, 4) = .008132$, which is not particularly good. Using inequality (5) we obtain: $B_{10}/P_{10} = .6790$ and $P(10 \leq x \leq 40) < .6790(.008132) = .005522$. Using (8) and calculating both B_{10} and B_{11}, we take $r = 11$ in the inequality (5) and obtain $B_{10} = .0035934$, $B_{11} = .0010889$, $P(11, 4) - P(40, 4) = .002840$, $B_{11}/P_{11} = .5657$, and hence $.004682 < P(10 \leq x \leq 40) < .003594 + .001607 = .00520$. Again using method (8), but calculating B_{12} also and using $r = 12$ in inequality (5), we get $.004974 < P(10 \leq x \leq 40) < .005099$, which is quite good. We can obtain a still better result by using inequality (7) instead of (5). Then $p' = 1/12$, $np' = 10/3$, $B_{10}/P'_{10} = 2.150+$, $P(10, 10/3) - P(40, 10/3) = .002366$, and $P(10 \leq x \leq 40) < .005087$.

ABSTRACTS OF PAPERS

Presented at the Madison Meeting of the Institute, September 7–10, 1948

1. On Distribution-free Confidence Intervals (Preliminary Report). WASSILY HOEFFDING, University of North Carolina, Chapel Hill.

Let $\theta(F)$ be a functional of a distribution function (d.f.) $F(x)$ (where x is a real number or a vector), defined over a class $\mathfrak{D}$ of d.f.'s; O_n a random sample from a population with d.f. $F(x)$; $\underline{\theta}_n \leq \bar{\theta}_n$ two functions of O_n; and $\alpha_n = Pr\{\underline{\theta}_n \leq \theta(F) \leq \bar{\theta}_n\}$. Conditions are studied under which, given α, $0 < \alpha < 1$, we have either $\alpha_n = \alpha$ or $\alpha_n \geq \alpha$ or $\alpha_n \to \alpha$, for all $F(x)$ in $\mathfrak{D}$, where $\mathfrak{D}$ is defined independently of the functional form of $F(x)$. Under fairly general conditions we can obtain by "studentization" confidence limits $\underline{\theta}_n$, $\bar{\theta}_n$ such that $\lim_{n\to\infty} \alpha_n = \alpha$, and $\gamma = \lim_{n\to\infty} E\sqrt{n}(\bar{\theta}_n - \underline{\theta}_n)$ exists; γ is minimized by using a least variance estimate of $\theta(F)$. If there exists a function $\kappa(\theta)$ such that var $T_n \leq \kappa^2(\theta)n^{-1}$ if $\theta(F) = \theta$, for all F in $\mathfrak{D}$, we can define confidence limits with a positive lower bound for α_n. This applies to a number of population characteristics estimated by rank order statistics, such as the coefficients ρ' and τ (estimated by Spearman's and Lindeberg-Kendall's rank correlation coefficients, respectively). In certain cases (including ρ' and τ), $\theta(F)$ admits a binomially distributed estimate; then exact confidence limits can easily be obtained. This research was done under an Office of Naval Research contract.

2. On Certain Statistics for Samples of 3 from a Normal Population. JULIUS LIEBLEIN, National Bureau of Standards, Washington.

In analytical chemistry three determinations are frequently made. Sometimes the average of only the two *closest* results is reported, the remaining observation being rejected as anomalous. In preparing a critique of this procedure, Dr. W. J. Youden encountered a need for information on certain properties of the distributions of the statistics $(x' - x'')/(x_3 - x_1)$, $(x' + x'')/2$, and $(x' - x'')/2$, where x' and x'' ($x' \geq x''$) are the two *closest* of the three determinations. This paper shows how these statistics differ from the ones heretofore treated involving "fixed" order statistics, gives the distribution of these statistics in random samples of 3 from a normal universe, and lists values of certain of the moments of their distributions.

3. On Multinomial Distributions with Limited Freedom: A Stochastic Genesis of Pareto's and Pearson's Curves. MARIA CASTELLAIN, University of Kansas City.

The purpose of this paper is to investigate the most probable configuration of N random elements to be distributed in $K(K < N)$ class intervals, where known forces are acting. We shall call these intervals of energy, using the terminology of statistical mechanics.

We will prove that the most probable configuration is a configuration of statistical equilibrium since its probability of occurring converges to 1 as N becomes infinitely large.

The main purpose of this paper is to discover which forces of attraction, operating in the intervals of energy, give Pareto's and Pearson's curves when statistical equilibrium is reached.

We will consider a random variable $Y(t)$, t being an independent variable, obeying a multinomial distribution law with limited freedom, and we will exploit the familiar process of statistical mechanics. The equation of the frequency curves corresponding to the equilibrium stage of the statistical experiment will be shown.

4. Fitting Generalized Truncated Normal Distributions. HAROLD HOTELLING, University of North Carolina, Chapel Hill.

In a sample from a p-dimensional normal distribution only those individuals are supposed to be observed which fall in a specified but arbitrary set A of positive measure. For estimating the parameters the method of moments is proved equivalent to that of maximum likelihood and therefore efficient. The problem is thus reduced to that of expressing the parameters of the normal distribution in terms of the moments of the truncated distribution. This however is not generally possible in simple explicit form. Methods are presented for dealing numerically with several special cases, including those in which A is a linear interval or a parallelogram.

5. On the Distribution of the Two Closest Observations Among a Set of Three Independent Observations. G. R. SETH, Iowa State College.

Let x_1, x_2, x_3 ($x_1 < x_2 < x_3$) be three independent ordered observations from a population having a probability density function $f(x)$. Let x', x'' ($x' < x''$) be the two closest, then the probability density function of x', x'' is given by

$$6 \cdot f(x') \cdot f(x'')[1 + F(2x'' - x') - F(2x' - x'')]$$

where

$$F(x) = \int_x^\infty f(x)\, dx.$$

In the case $f(x)$ is a normal distribution with unit variance, the joint distribution of $y = x'' - x'$ and $z = \dfrac{x'' - x'}{x_3 - x_1}$ is obtained as

$$\frac{2\sqrt{3y^2}}{\pi z^2} \exp\left[-\frac{y^2(1 - z + z^2)}{3z^2}\right].$$

This problem is of interest in cases where the conclusions are to be based on a set of three observations and one of the observations is to be rejected in the analysis of the data

6. The Derivation of Certain Recurrence Formulae and their Application to the Extension of Existing Published Incomplete Beta Function Tables. T. A. BANCROFT, Alabama Polytechnic Institute, Auburn (presented by title).

The objects of the paper are: (1) to give a number of new recurrence formulae in the incomplete beta function derived by a new method, and (2) to indicate how these new formulae have been used to obtain new tables of the incomplete beta function that are outside the range of the p and q values given in the existing published tables.

The recurrence formulae have been derived by considering the incomplete beta function as a special case of the hypergeometric series, thus

$$B_x(p, q) = \frac{x^p}{p} F(p, 1 - q, p + 1, x),$$

where the usual form of the hypergeometric series is

$$F(a, b, c, x) = 1 + \frac{a \cdot b}{c} \frac{x}{1!} + \frac{a(a+1) \cdot b(b+1)}{c(c+1)} \frac{x^2}{2!} + \frac{a(a+1)(a+2) \cdot b(b+1)(b+2)}{c(c+1)(c+2)} \frac{x^3}{3!} + \cdots.$$

This series converges for $|x| < 1$, and $x = 1$, if and only if $a + b < c$. Certain recurrence formulae for $F(a, b, c, x)$ are then directly converted for use with $B_x(p, q)$, or in the so-called normalized form $I_x(p, q)$, provided $c = a + 1$. All conditions have been satisfied by setting $a = p, b = 1 - q, c = p + 1$, and $q > 0$.

For example, using the above mentioned methods we may obtain, among many others, the recurrence formulae

(i) $xI_x(p, q) - I_x(p + 1, q) + (1 - x)I_x(p + 1, q - 1) = 0$,

(ii) $(p + q - px)I_x(p, q) - qI_x(p, q + 1) - p(1 - x)I_x(p + 1, q - 1) = 0$,

(iii) $qI_x(p, q + 1) + pI_x(p + 1, q) - (p + q)I_x(p, q) = 0$.

Formula (i) is essentially the basic recurrence formula used to obtain Karl Pearson's tables An indication of formula (iii) in another form was given by the author in the paper "On Biases in Estimation Due to the Use of Preliminary Tests of Significance," *Annals of Math. Stat.*, Vol. 15 (1944), p. 194, and a direct proof was later given by the author in "Note on an Identity in the Incomplete Beta Function," *Annals of Math Stat.* Vol. 16 (1945), pp. 98–99. All of the material in the present paper, however, is new, including recurrence formulae and tables and the mathematical method of derivation.

7. Asymptotic Studentization in Testing of Hypotheses. HERMAN CHERNOFF, Cowles Commission for Research in Economics.

If H is a hypothesis for which $t \leq c_1(\theta)$ would be a good test if the value of the nuisance parameter θ were known and $\hat{\theta}$ is an estimate of θ, then the following method of asymptotic studentization (obtaining critical regions of almost constant size) was suggested by Wald Consider $t \leq \varphi(\hat{\theta})$ where $\varphi(\hat{\theta}) = c_1(\hat{\theta}) + \cdots + c_s(\hat{\theta})$ and $Pr\{t \leq c_1(\theta)\} = \alpha$, $Pr\{t - c_1(\hat{\theta}) \leq c_2(\theta)\} = \alpha, \cdots Pr\{t - c_1(\hat{\theta}) - \cdots - c_r(\hat{\theta}) \leq c_{r+1}(\theta)\} = \alpha$. It is shown that under reasonable conditions this test, and various modifications, designed for those cases where the $c_r(\theta)$ are difficult to obtain exactly have the asymptotic property that $Pr\{t \leq \varphi(\hat{\theta})\} = \alpha + O(N^{-s/2})$ where N is the size of the sample involved or an analogous variable This property can be extended to the case where θ is a k-dimensional variable.

8. Completeness, Similar Regions, and Unbiased Estimation. (Preliminary Report.) ERICH L. LEHMANN AND HENRY SCHEFFÉ, University of California at Los Angeles.

A family $\mathfrak{M}$ of measures M on a space X of points x is defined to be *complete* if $\int_X f(x)\, dM = 0$ for every M in $\mathfrak{M}$ implies $f(x) = 0$ except on a set A for which $M(A) = 0$ for every M in $\mathfrak{M}$. For a given family of measures the question of completeness may be regarded as the question of unicity of a related functional transform. Classical unicity results are applicable to many families of probability distributions that have been studied by statisticians. The notion of completeness throws light on the problem of similar regions and the problem of unbiased estimation. The concept of a *maximal* sufficient statistic—roughly, a sufficient statistic that is a function of all other sufficient statistics—is developed A constructive method of finding such is given, which seems to apply to all examples ordinarily considered in statistical theory. A relation between completeness and maximality is found.

9. On a Proposed Method for Estimating Populations. CECIL C. CRAIG, University of Michigan, Ann Arbor.

It was proposed to the author by a biologist that a method be devised for estimating the total population in an area which shall utilize the minimum distances between randomly

chosen individuals and their neighbors in directions lying in each of the four quadrants. Assuming that the area is a square and that the distribution law over it is rectangular, it turns out that the complete distribution of the lengths of sides of minimum squares which contain a second individual is simpler than that of minimum distances. In both cases a simple estimate is found which uses most but not all of the information in a sample and whose efficiency is comparable to that based on a complete enumeration of a sample area, though such an enumeration is not always possible.

10. Some Results on the Asymptotic Distribution of Maximum- and Quasi-Maximum-likelihood Estimates. Herman Rubin, Institute for Advanced Study.

The author investigates the asymptotic normality of maximum- and quasi-maximum-likelihood estimates of parameters of systems of linear stochastic difference equations The principal tool is the extension of the Central Limit Theorem to dependent variables previously obtained by the author (presented to the American Mathematical Society in April, 1948). The results obtained are analogous to those in the case in which no differences are present. Some extensions are also made to systems of stochastic difference equations linear in the coefficients but not necessarily in the variables. If the complete system of stochastic difference equations is linear in the jointly dependent variables, asymptotic efficiency is demonstrated for maximum-likelihood estimates.

11. The Probability Points of the Distribution of the Median in Random Samples from Any Continuous Population. Churchill Eisenhart, Lola S. Deming, and Celia S. Martin, National Bureau of Standards, Washington.

The abscissa of the (one-tail) ϵ-probability point of the distribution of the median in random samples of size $n = 2m + 1$ $(m \geq 0)$ from any continuous population is identical with the abscissa of the corresponding $P_{\epsilon,n}$-probability point of the parent distribution, where $P_{\epsilon,n}$ is determined by

$$(1) \qquad \sum_{k=\frac{1}{2}(n+1)}^{n} C_k^n P_{\epsilon,n}^k (1 - P_{\epsilon,n})^{n-k} = \epsilon, \qquad (0 \leq \epsilon \leq 1).$$

From (1) it follows that

$$(2) \qquad P_{1-\epsilon,n} = 1 - P_{\epsilon,n}$$

and that

$$(3) \qquad P_{\epsilon,n} = x_\epsilon(n+1, n+1) = \frac{1}{1 + F_\epsilon(n+1, n+1)} = \frac{1}{1 + e^{2Z_\epsilon(n+1,n+1)}},$$

where $x_\epsilon(\nu_1, \nu_2)$, $F(\nu_1, \nu_2)$, and $Z_\epsilon(\nu_1, \nu_2)$ denote the ϵ-probability points of the incomplete-beta-function distribution, Snedecor's F-distribution and Fisher's z-distribution, for $\nu_1(= 2q)$ and $\nu_2(= 2p)$ 'degrees of freedom', respectively. The foregoing results are certainly not "new": Harry S. Pollard implicitly utilized the first equality on the extreme left of (3) in his doctoral dissertation at the University of Wisconsin in 1933 (see *Annals of Math Stat.*, Vol. 5 (1934), p. 250), and John H. Curtiss has given the generalization of (1) appropriate to the case of the 'rth. position' in random samples from any continuous population (see *Amer. Math. Monthly*, Vol. 50 (1943), p. 103) and utilized (3) explicitly to obtain the 5% point of the distribution of the median in random samples of size $n = 23$. The aim of the present paper is to give these results somewhat greater publicity—they are hardly "well known". To this end a table (Table 1) is given of the values of $P_{\epsilon,n}$ to 5 significant figures for $\epsilon = 0.001, 0.005, 0.01, 0.025, 0.05, 0.10, 0.20, 0.25$ and $n = 3(2)15(10)95$, together

with expressions from which $P_{\epsilon,n}$ can be evaluated accurately and conveniently for values of n (and ϵ) not included in the table. Numerical examples illustrate the use of the table and formulas. Concise derivations of the fundamental relations and formulas are given in an appendix.

12. On the Arithmetic Mean and the Median in Small Samples from the Normal and Certain Non-Normal Populations. CHURCHILL EISENHART, LOLA S. DEMING, and CELIA S. MARTIN, National Bureau of Standards, Washington.

Let $\bar{x}_{\epsilon,n}$ and $\tilde{x}_{\epsilon,n}$ denote the abscissae of the one-tail ϵ-probability points of the arithmetic mean and the median, more specifically, the abscissae *exceeded* with probability ϵ by the mean and the median, respectively, in random samples of size n $(= 2m + 1)$ from any specified population, and let $\sigma_{\bar{x}_n}$ and $\sigma_{\tilde{x}_n}$ denote the standard deviations of the mean and the median in such samples, respectively. The following symmetrical populations with zero location parameters and unit scale parameters are considered in this paper

Type		
normal (Gaussian)	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$,	$-\infty \leq x \leq \infty$
double-exponential (Laplace)	$\frac{1}{2}e^{-\lvert x \rvert}$,	$-\infty \leq x \leq \infty$
rectangular (uniform)	1,	$-\frac{1}{2} \leq x \leq \frac{1}{2}$
Cauchy	$\frac{1}{\pi}\frac{1}{1+x^2}$,	$-\infty \leq x \leq \infty$
sech	$\frac{1}{\pi}\operatorname{sech} x$,	$-\infty \leq x \leq \infty$
sech² (derivative of "logistic")	$\frac{1}{2}\operatorname{sech}^2 x$,	$-\infty \leq x \leq \infty$

Using the basic table, relating probability points of the distribution of the median to probability points of the parent distribution, given in Churchill Eisenhart, Lola S Deming and Celia S. Martin, "The probability points of the distribution of the median in random samples from any continuous population," values of $\tilde{x}_{\epsilon,n}$ for random samples from each of the above distributions have been evaluated, and are tabulated to 5 decimal places in the pres- 0 001, 0.005, 0.01, 0 025, 0 05, 0 10, 0.20, 0.25.
ent paper, for $n = 3(2)15(10)95$ and $\epsilon =$

In the case of the *normal distribution*, values of $\bar{x}_{\epsilon,n}$ to 5 decimal places are given also for the aforementioned combinations of ϵ and n Comparison of the values of $\bar{x}_{\epsilon,n}$ and $\tilde{x}_{\epsilon,n}$ gives precise numerical meaning to the well-known lesser accuracy of the median as an estimator of the center of a normal population, for samples of *any* odd size $(n = 2m + 1)$. Values of the ratio $R_{\epsilon,n} = \tilde{x}_{\epsilon,n}/\bar{x}_{\epsilon,n}$ are given also for this case (normal population), to 4 decimal places for the above combinations of ϵ and n, together with the best available values of $\sigma_{\tilde{x}_n}/\sigma_{\bar{x}_n}$ for $n = 3(2)15(10)55$. When $0 < \epsilon \leq 0.025$, the ratio $R_{\epsilon,n}$ exceeds the ratio $\sigma_{\tilde{x}_n}/\sigma_{\bar{x}_n}$, showing that the 'tails' of the exact distribution of the median are 'longer' than the tails of the normal distribution with the same mean and standard deviation, and, when $0.05 \leq \epsilon \leq 0.25$, the ratio $R_{\epsilon,n}$ is less than $\sigma_{\tilde{x}_n}/\sigma_{\bar{x}_n}$. (A theoretical argument shows that the point of equality is close to the 0.042-probability point.) A method for computing $\sigma_{\tilde{x}_n}$, based on the foregoing, is given that is believed to be accurate to $.001/\sqrt{n}$, or better for $n \geq 3$.

In the case of the *double-exponential* distribution, values of $\bar{x}_{\epsilon,n}$ are given to 4 decimal places for $n = 3(2)11$, and $\epsilon = 0.005, 0.01, 0\ 025, 0.05, 0\ 10, 0\ 25$, for comparison with the corresponding values of $\tilde{x}_{\epsilon,n}$. It is found that *when* $n = 3$, $\bar{x}_{\epsilon,3} < \tilde{x}_{\epsilon,3}$ for $\epsilon = 0\ 005, 0.001$, and 0.025, indicating that in random samples of 3 from a double-exponential distribution the arithmetic mean furnishes narrower confidence limits for the center of the distribution

at 0.95, 0.98, and 0.99 levels of confidence. When $n = 5$, the mean is 'better' at the .98 and .99 levels of confidence; and, when $n = 7$, at the 0.99 level. For all other combinations of ϵ and n (≥ 3), the median is 'better.'

In the case of the *rectangular* distribution, values of $\bar{x}_{\epsilon,n}$ are tabulated to 4 decimals for $n = 3(2)9$, and values of $\tilde{x}_{\epsilon,n}$, the ϵ-probability point of the mid-range in samples of n, for $n = 3(2)15(10)95$, in each instance for $\epsilon = 0.005, 0.01, 0.025, 0.05, 0.10, 0.25$, and in the case of $\bar{x}_{\epsilon,n}$ for $\epsilon = 0.001$ also. The superiority of the midrange over the mean and the median, well-known but here exhibited numerically for the first time, is truly amazing.

It is planned to provide values of $\bar{x}_{\epsilon,n}$ for samples from the sech and sech² distributions in the final paper.

13. The Relative Frequencies with which Certain Estimators of the Standard Deviation of a Normal Population Tend to Underestimate its Value. CHURCHILL EISENHART and CELIA S. MARTIN, National Bureau of Standards, Washington.

Let $x_1, x_2, \cdots, x_n$ denote a random sample of n independent observations from a normal population with mean μ and standard deviation σ. Common estimators of σ are

$$s_1 = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2/n}, \qquad s_2 = s_1\sqrt{n/(n-1)}, \qquad s_3 = s_1/c_2,$$

$$m_1 = \sqrt{\frac{\pi}{2}}\sum_{i=1}^{n} |x_i - \bar{x}|/n, \qquad m_2 = m_1\sqrt{n/(n-1)},$$

and $R_1 = (x_L - x_S)/d_2$, where $\bar{x} = \sum_{i=1}^{n} x_i/n$, x_L is the largest and x_S the smallest of the x's, $c_2 = \mathrm{E}(s_1)$, and $d_2 = \mathrm{E}(x_L - x_S)$, the symbol E( ) denoting "mathematical expectation (or mean value) of." A table is given that shows to 3 decimals the relative frequencies (probabilities) with which these estimators tend to underestimate σ when $n = 2(1)10$, 12, 15, 20, 24, 30, 40, 60. The results show among other things that, for very small samples ($n \leq 10$) such as chemists and physicists commonly use, Bessel's formula for the probable error, which is based on s_2, has a marked downward bias in the probability sense (in addition to its known slight downward bias in the mean value sense), whereas Peter's formula, which is based on m_2, has only a slight downward bias in the probability sense and no bias in the mean value sense. A table of divisors is given by means of which "median estimators" of σ can be computed readily from the basic quantities $\sum_{i=1}^{n} (x_i - \bar{x})$, $\sum_{i=1}^{n} |x_i - \bar{x}|$, and $(x_L - x_S)$, that is, estimators that will over- and underestimate σ equally often in repeated use. An application to control charts is noted. Median estimators, like maximum likelihood estimators ("modal estimators") have the useful property that if $T_{\frac{1}{2}}$ is a median estimator of θ, then $f(T_{\frac{1}{2}})$ is a median estimator of $f(\theta)$, a property unfortunately not possessed by the customary "unbiased" ("mean") estimators.

14. Some Non-Parametric Tests of Whether the Largest Observations of a Set are too Large. (Preliminary Report.) JOHN E. WALSH, Douglas Aircraft Company, Santa Monica, California.

Let $x(1), \cdots, x(n)$ represent the values of n observations arranged in increasing order of magnitude. By hypothesis these observations have the properties: (1) They are independent and from continuous symmetrical populations (2) For large n the variances of the tail order statistics are either very large or very small compared with the variances of the central order statistics (3) For large n the tail order statistics are approximately independent

of the central order statistics (4) Each observation is from a population whose median is either θ or φ, where $x(n - i + 1), \cdots, x(n)$ are from populations with median θ while the central and smaller order statistics are from populations with median φ. The test is: *Accept* $\varphi < \theta$ *if* $\min [x(n - i_k) + x(j_k), 1 \leq k \leq s \leq r] > 2x(t_\alpha)$, where $i_u < i_{u+1}$, $j_v < j_{v+1}$, $i_s = r - 1$, and t_α is defined by $Pr\, [x(t_\alpha) < \varphi \mid \theta = \varphi] = \alpha$. Here

$$\alpha = \Pr \{\min [x(n - i_k) + x(j_k); 1 \leq k \leq s \leq r] > 2\varphi \mid \theta = \varphi\}.$$

For large n the significance level of the test is approximately α while the significance level does not exceed $2\,\alpha$ for any value of n Suitable values of α can be obtained for $r \geq 4$. As $\theta - \varphi \to -\infty$ the power function tends to zero, while the power function tends to unity as $\theta - \varphi \to \infty$. For $\theta - \varphi < 0$ the power function is monotonically increasing

15. On the Bounded Significance Level Properties of the Equal-tail Sign Test for the Mean. John E. Walsh, Douglas Aircraft Company, Santa Monica, California, (Presented by Title).

The equal-tail sign test for deciding whether the population mean μ is equal to a given hypothetical value μ_0 is defined by: *Accept* $\mu \neq \mu_0$ *if either* $x_i < \mu_0$ *or* $x_{n+1-i} > \mu_0$, $\left(i > \dfrac{n+1}{2}\right)$ Here x_j, $(j = 1, \cdots, n)$, is the j^{th} largest of n independent observations drawn from n populations which satisfy the conditions. (i) The mean of each population has the value μ. (ii) Each population is continuous at its mean (iii) The mean is at a 50% point for each population. This paper investigates how the significance level of the equal-tail sign test varies when (i)–(iii) are not satisfied. It is found that the significance level does not differ noticeably from its hypothetical value under conditions much more general than (i)–(iii). This significance level stability, combined with the properties of being easily applied and reasonably efficient for small samples from a normal population, suggests that the equal-tail sign test be considered for application whenever the population mean is to be tested on the basis of a small number of observations

16. Infinitely Divisible Distributions. William Feller, Cornell University, Ithaca, New York.

A simple derivation of P. Lévy's formula is given starting from the following definition: a distribution function $F(x)$ is infinitely divisible if for every n it is possible to find finitely many distributions $F_{k,n}(x)$ such that $F(x) = F_{1,n}(x) * \cdots * F_{k_n,n}(x)$ and that $F_{k,n}(x)$ tends to the unitary distribution uniformly in n. This definition is more general than the one used by P. Lévy and Khintchine. The equivalence of the two definitions was proved by Khintchine by deep methods The new approach renders the equivalence obvious Furthermore, a new characterization of infinitely divisible distributions is given; it is equivalent to Gnedenko's characterization but requires no special analytical tools.

17. Fluctuation Theory of Recurrent Events. William Feller, Cornell University, Ithaca, New York.

Consider a sequence of independent or dependent trials but suppose that each has a discrete sample space. The paper studies recurrent patterns $\mathfrak{E}$ which can be roughly characterized by the property that after every occurrence of $\mathfrak{E}$ the process starts from scratch, the conditional probabilities coinciding with the original absolute probabilities Typical examples are success runs, returns to equilibrium, zeros of sums of independent variables, passages through a state in a Markov chain. New methods are developed unifying and simplifying previous theories and applying to larger classes of recurrent events It is shown

in an elementary way the probability that $\mathfrak{E}$ occurs at the n-th trial either has a limit or is asymptotically periodic. This theorem has many consequences For example, the ergodic properties of discrete Markov chains follow in a few lines, and the difference between finite and infinite chains disappears. Several theorems of the renewal type are proved Weak and strong limit theorems for the number N_n of occurrences of $\mathfrak{E}$ in n trials are derived shedding new light on stable distributions.

18. Formulas for the Percentage Points of the Distributions of the Arithmetic Mean in Random Samples from Certain Symmetrical Universes. UTTAM CHAND, University of North Carolina and National Bureau of Standards.

Using the method of Fisher and Cornish, the $100\epsilon\%$ point of the distribution of the arithmetic mean in random samples of size N from any universe having finite cumulants of the first four orders, κ_1, κ_2, κ_3, κ_4, is expressed to order $1/N^2$ as a function of N, the $100\epsilon\%$ point of a standardized normal deviate and the quantities κ_1, κ_2, $\kappa_3/\kappa_2^{3/2}$, κ_4/κ_2^2 The numerical coefficients are evaluated for the cases of sampling from rectangular, double-exponential, sech and sech^2 distributions. The application of the resulting formulas is illustrated numerically for ϵ = .001, .005, .010, .025, .050, .100, and .250. In the case of the rectangular and double-exponential distributions, the results obtained for $N = 10$ are compared with accurate values, indicating the accuracy of the formulas.

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Professor T. A. Bickerstaff has been appointed Chairman of the Department of Mathematics at the University of Mississippi.

Professor Raj Chandra Bose has resigned as head of the graduate Department of Statistics of the University of Calcutta, and has been appointed Professor of Mathematical Statistics at the University of North Carolina beginning in the winter of 1949. Professor Bose is an authority on the design of experiments and is writing a book on the combinatorial mathematics of the subject. He has also published extensive contributions to differential geometry and to multivariate statistical analysis, and has been instrumental in developing practical sample surveys. He served as Visiting Professor in the Institute of Statistics at North Carolina in the winter and spring of 1948

Mr. Hamilton Brooks's paper, "The Probable Breakdown Voltage of Paper Dielectric Capacitors," was one of the four papers selected for a national award by the American Institute of Electrical Engineers. His paper presents the statistical treatment of an engineering problem and shows by experiment how insulation strength distribution is determined by the distribution of the extreme size of flaws.

Dr. C. West Churchman, formerly a member of the staff at the University of Pennsylvania, was appointed Associate Professor of Philosophy at Wayne University, Detroit 1, Michigan, starting February 1, 1948.

Dr. William G. Cochran has accepted an appointment as Professor of Biostatistics in the School of Hygiene and Public Health of the Johns Hopkins University and will assume this post in September. Dr. Cochran, a native of Glasgow, Scotland, comes to Johns Hopkins from the University of North Carolina where he served as Associate Director of the Institute of Statistics from 1946 until the present.

Dr Louis M. Court has been promoted to an assistant professorship in the Mathematics Department of Rutgers University.

Dr. Donald A. Darling, formerly a member of the staff at Cornell University, has accepted an assistant professorship at Rutgers University.

Mr. Aryeh Dvoretzky has been appointed a member of the Institute for Advanced Study, Princeton, New Jersey, for the 1948–1949 academic year.

Mr. Arnold King, formerly Director of Research in Statistical Methodology for the Bureau of Agricultural Economics at Iowa State College, was appointed Managing Director of National Analysts, Inc., Philadelphia on July 1, 1948.

Mr Charles L. Marks has resigned his position as instructor of mathematics at the University of North Carolina to accept a teaching appointment in the Department of Statistics, The George Washington University, Washington 6, D. C.

Miss Doris Newman has accepted an appointment at the U. S. Naval Medical Research Laboratory, U. S. Naval Submarine Base, New London, Conn.

Dr. Ernest Rubin has been transferred from the Immigration and Naturalization Service, General Research Section, Washington, D. C. to the European Branch, Areas Division, Office of International Trade in the Department of Commerce as an Economic Statistician.

Mr. David Rubinstein has been promoted from Junior Research Assistant in the Statistical Laboratory, University of California, Berkeley, to a Teaching Assistant.

Miss Elizabeth L. Scott, formerly an Associate and Research Assistant in the Statistical Laboratory, University of California, Berkeley, has been promoted to Lecturer and Research Assistant.

Dr. Gobind R. Seth, who was formerly a student at Columbia University, has accepted an associate professorship in statistics at the Statistical Laboratory, Iowa State College.

Dr. Charles M. Stein has been promoted to an assistant professorship in the Statistical Laboratory, University of California, Berkeley.

Professor Gerhard Tintner is on leave of absence for one year from the Iowa State College to join the Department of Applied Economics at Cambridge University, Cambridge, England as a Research Associate.

Mr. L. H. C. Tippett, Chief Statistician of the British Cotton Industry Research Association, delivered twelve one-hour lectures on *Statistical Quality Control and Industrial Experimentation* at a conference at the Massachusetts Institute of Technology, May 5–14, before a large audience. Dr. W. A. Shewhart of the Bell Telephone Laboratories addressed a large audience on the *Future of Statistics in Industrial Research and Quality Control* on May 14 at the same conference

Scientists and Reserve Officers

The Department of the Army has established a program of particular interest to statisticians and other scientists who hold Reserve commissions in the Army, and who are professionally engaged in teaching or research and development.

The objectives of the program are to:

(1) maintain the useful affiliation of statisticians and other scientists with the Organized Reserve Corps,

(2) provide peacetime Reserve assignments for these officers, enabling optimum utilization of their education, experience and skills,

(3) furnish mobilization assignments which will fully utilize their talents, and

(4) adequately prepare these officers for mobilization.

The Technical Services of the Department of the Army submit to these Research and Development Reserve Groups research problems and projects which pose an intellectual challenge to members of the group. Thus, the program provides members of each group a type of training which is in keeping with their scientific and technical interests and competence, rather than a traditional kind of training session in which scientists have little or no interest.

The program is now being implemented only in those areas where there is a definite local interest To date, eighteen Research and Development Reserve groups have been organized. Twelve additional groups are in process of organization. Others are in the initial stages of formation Several of these groups have been formed in communities in which large universities, industrial research laboratories, or private research foundations are located Typical localities are Chicago, Illinois, Wilmington, Delaware, Newark, New Jersey; Houston, Texas; Washington, D C.; Manhattan and Lawrence, Kansas, Champaign-Urbana, Illinois; Pittsburgh, Pennsylvania, Denver, Colorado; and Detroit, Michigan

Provision is made to submit research projects of interest to all categories of scientists—chemists, physicists, engineers, geologists, geographers, psychologists, mathematicians, statisticians and all of the biological scientists.

Reserve officers who are currently engaged in civilian research, college or university teaching, or industrial research or development, or who in the past have had specific research experience are eligible to make application for assignment to an Organized Reserve Research and Development Group. A group may be organized in any locality where there are twenty (20) or more qualified officer scientists who desire to participate in the program A subgroup may be organized with ten (10) qualified members

The program is under the general direction of the Research and Development Group, Logistics Division, General Staff, United States Army. The entire program is outlined in Department of the Army Circular Number 127, dated 5 May 1948.

Inquiry about organization of an Organized Reserve Research and Development Group or about assignment to a group already organized should be made of the Unit Instructor, ORC, or of the Senior Army Instructor, ORC, in the locality in which the officer resides In localities in which a group has already been organized, the Commanding Officer of the group will consider applications for assignment of additional officers

New Members

The following persons have been elected to membership in the Institute

(June 1 to August 15, 1948)

Anderson, Hjalmar, Jr. (Univ. of Oregon Medical School) Student, *Turner, Oregon.*

Banerjee, Kali Shankar, M.A. (Calcutta Univ) Statistician, Central Sugar Cane Research Station, P.O. Pusa, Bihar, India.

Bordelow, Derrill Joseph, B.S. (Louisiana State Univ.) Associate Physicist with Naval Ordnance Laboratory, *602 A. Street S E., Washington, D. C.*

Cowan, David, B S (Tufts Univ.) Research Analyst, War Department, *89 Lewis Street, East Lynn, Massachusetts*

Frederiksen, Norman, Ph D (Syracuse Univ.) Research Associate, Educational Testing Service, Associate Professor of Psychology, Princeton University, Educational Testing Service, Box 592, Princeton, N. J

Gehman, Harry M., Ph.D. (Univ. of Pennsylvania) Professor of Mathematics, University of Buffalo, *163 Winspear Avenue, Buffalo 15, New York*

Hofmann, John E., A.M. (Univ. of Minnesota) Senior Research Fellow, *3222 Oakland Street, Ames, Iowa*

Kimball, Allyn W., Jr., B.S. (Univ. of Buffalo) Research Statistician, Department of Biometrics, School of Aviation Medicine, Randolph Field, Texas

King, Edgar P., Jr., B.S. (Carnegie Institute of Technology) Teaching Assistant in Mathematics, Department of Mathematics, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania

Link, Curtis K., B.S. (Univ. of Oregon) Graduate Student-Assistant, *750 W. 6th Street, Eugene, Oregon*

Leider, Nathan, B.A. (College of the City of N. Y.) Mathematician P-2, *1841 Summit Place, N.W., Washington 9, D. C.*

Manos, Nicholas E., M.A. (Univ. of Calif.) Meteorologist and Statistician, *1424 Rhode Island Avenue, N.W., Washington 5, D. C.*

Peters, Stefan, Ph.D. (Erlangen, Germany) Lecturer at the University of California, *1207 Peralta Avenue, Berkeley 6, California.*

Petrou, Nicholas V., M.Sc. (Harvard Univ.) Electrical Engineer, Project Engineer, Westinghouse Electric Corporation, *1844 Ardmore Blvd., Pittsburgh 21, Pennsylvania.*

Prakash, Aditya, M.A. (Univ. of Michigan) Student, c/o Mathematics Department, University of Michigan, Ann Arbor, Michigan.

Read, Robert R., B.S. (Oregon State College) Apprentice Engineer, Inventory and Costs Division, Pacific Telephone and Telegraph Company, *3207 N.E., 30, Portland, Oregon*

Seiden, Esther, M.A. (Vilno, Poland) Research Assistant, Statistical Laboratory, University of California, *2116 Derby Street, Berkeley 5, California.*

Sodano, John J., B.S. (Queens College) Student, Mathematical Statistics, Columbia University, *172-15 93rd Avenue, Jamaica 3, New York.*

Stillinger, Richard C., M.S., (Univ. of Michigan) Graduate Student, *1808 Weston Court, Willow Run, Michigan.*

Swan, Albert W., B.A.Sc. (Univ. of Toronto) Statistical Section Research and Development Department, The United Steel Company Limited, c/o The United Steel Companies Ltd., 17 Westbourne Road, Sheffield 10, England.

Tate, Robert F., A.B. (Univ. of Calif.) Teaching Fellow, Department of Mathematical Statistics, *Phillips Hall, Chapel Hill, North Carolina.*

Teichroew, Dan, B.A. (Univ. of Toronto) Division of Research, Department of Lands and Forest, South Baymouth, Ontario, Canada.

Tyler, Leona E., Ph.D. (Univ. of Minnesota) Associate Professor of Psychology, Department of Psychology, University of Oregon, Eugene, Oregon.

ADOPTION OF THE NEW CONSTITUTION

The chief order of business at the business meeting of the Institute held at Madison, Wisconsin on September 10, 1948, was the adoption of the new Constitution. The draft mailed to the members in August, 1948, was adopted unanimously after two changes had been made. They were: (1) the insertion of the word "Article" before each of the respective articles and (2) the elimination of the first "the" in the third line and fourth paragraph of Article 4.

Other business transacted at the meeting included a report of the Secretary-Treasurer on the financial condition of the Institute indicating that while the Institute is just operating within its income during 1948, steps will have to be taken to provide the additional revenue needed for 1949. It was decided not to raise dues for 1949 but to attempt to raise additional funds by: (1) an immediate appeal to universities and other institutions which are sponsoring research in mathematical statistics for contributions to the Institute and (2) an appeal to the members of the Institute to make additional contributions at the time of the payment of their annual dues.

Other matters under consideration at the meeting included a reading and discussion of a proposed revision of the By Laws, the announcement of the dates and locations of future meetings of the Institute and the passing of a resolution of thanks to those contributing to the success of the Madison meeting.

A copy of the official minutes of this meeting may be obtained on request from the Secretary-Treasurer.

P. S. Dwyer
Secretary-Treasurer

REPORT ON THE MADISON MEETING OF THE INSTITUTE

The Eleventh Summer Meeting of the Institute of Mathematical Statistics was held at the University of Wisconsin, Madison, Wisconsin, Tuesday, September 7 through Friday, September 10, 1948. The meeting was held in conjunction with the summer meetings of the American Mathematical Society, the Mathematical Association of America and the Econometric Society. The following eighty members of the Institute attended the meeting:

C. B. Allendoerfer, V. L. Anderson, K. J. Arnold, H. M. Bacon, A. S Barr, Walter Bartky, H. P. Beard, A. A. Bennett, T. A. Bickerstaff, J. H. Bushey, Maria Castellani, Uttam Chand, Herman Chernoff, C. C Craig, J. H. Curtiss, G. B Dantzig, D. B. De Lury, J. L Doob, A M. Dutton, P. S. Dwyer, Mrs. Daisy Edwards, Churchill Eisenhart, H. P. Evans, C. H. Fischer, J. E. Freund, H. M. Gehman, H. H. Germond, M. A. Girshick, Casper Goffman, P. R Halmos, W. G. Hart, E. H. C. Hildebrandt, Wassily Hoeffding, D. G Horvitz, Harold Hotelling, A. S. Householder, M. H. Ingraham, Leo Katz, Oscar Kempthorne, J. F Kenney, W. M. Kincaid, T. C. Koopmans, H. D. Larsen, Walter Leighton, H. B. Mann, A. M. Mark, Jacob Marschak, A. W. Marshall, Kenneth May, M. R. Mickey, Jr., Dorothy J. Morrow, C J. Nesbitt, M. J Netzorg, John von Neumann, Jerzy Neyman, G. B. Price, C. J. Rees, J. S. Rhodes, P R. Rider, F. D. Rigby, Herman Rubin, Arthur Sard, Henry Scheffé, E D. Schell, I. E Segal, G. R. Seth, W. B. Simpson, Andrew Sobczyk, E W. Stacy, C. M. Stein, A. G. Swanson, Zenon Szatrowski, R. M. Thrall, A. W Tucker, J. W. Tukey, W. A. Wallis, J. E. Walsh, J. E. Wilkins, Jr., S. S. Wilks, M. A. Woodbury.

The Tuesday morning session was devoted to contributed papers. Professor K. J. Arnold of the University of Wisconsin presided. The attendance was approximately forty. The following papers were presented:

1. *On Distribution-free Confidence Intervals.* Preliminary Report.
   Dr Wassily Hoeffding, Institute of Statistics, University of North Carolina.
2. *On Certain Statistics for Samples of 3 from a Normal Population.*
   Mr Julius Lieblein, Statistical Engineering Laboratory, National Bureau of Standards Presented by Dr. Churchill Eisenhart.
3. *On Multinomial Distributions with Limited Freedom: A Stochastic Genesis of Pareto's and Pearson's Curves.*
   Professor Maria Castellani, University of Kansas City.
4. *Fitting Generalized Truncated Normal Distributions*
   Professor Harold Hotelling, Institute of Statistics, University of North Carolina.
5. *On the Distribution of the Two Closest Observations Among a Set of Three Independent Observations.*
   Professor G. R. Seth, Statistical Laboratory, Iowa State College.
6. *The Derivation of Certain Recurrence Formulae and their Application to the Extension of Existing Published Incomplete Beta Function Tables.*
   Dr. T. A. Bancroft, Alabama Polytechnic Institute. (Presented by title)

On Tuesday afternoon a session for contributed papers was held jointly with the American Mathematical Society. Professor P. S. Dwyer of the University of Michigan presided The attendance was approximately eighty. The following papers were presented:

7. *Asymptotic Studentization in Testing Hypothesis*
   Dr Herman Chernoff, Cowles Commission, University of Chicago.
8. *Completeness, Similar Regions and Unbiased Estimation* Preliminary Report
   Professor E. L Lehman, University of California and Professor Henry Sheffé, University of California at Los Angeles
9. *On a Proposed Method for Estimating Populations*
   Professor C. C Craig, University of Michigan
10. *Some Results on the Asymptotic Distribution of Maximum- and Quasi-maximum-likelihood Estimates*
    Dr Herman Rubin, Institute for Advanced Study.
11. *The Probability Points of the Distribution of the Median in Random Samples from any Continuous Population.*
    Dr. Churchill Eisenhart, Lola S Deming and Celia S Martin, Statistical Engineering Laboratory, National Bureau of Standards
12. *On the Arithmetic Mean and the Median in Small Samples from the Normal and Certain Non-normal Populations.*
    Dr Churchill Eisenhart, Lola S Deming and Celia S Martin, Statistical Engineering Laboratory, National Bureau of Standards.
13. *The Relative Frequencies with which Certain Estimators of the Standard Deviation of a Normal Population Tend to Underestimate Its Value*
    Dr Churchill Eisenhart and Celia S Martin, Statistical Engineering Laboratory, National Bureau of Standards
14. *Some Non-parametric Tests of Whether the Largest Observations of a Set are too Large* Preliminary Report
    Dr J. E Walsh, Project Rand, Santa Monica, California
15. *On Some Bounded Significance Level Properties of the Equaltail Sign Test for the Mean.*
    Dr. J E. Walsh, Project Rand, Santa Monica, California. (Presented by title.)
16. *Infinitely Divisible Distributions.*
    Professor Will Feller, Cornell University (Presented by title.)
17. *Fluctuation Theory of Recurrent Events*
    Professor Will Feller, Cornell University (Presented by title)
18. *Formulae for the Percentage Points of the Distributions of the Arithmetic Mean in Random Samples from Certain Symmetrical Universes*
    Mr. Uttam Chand, University of North Carolina and National Bureau of Standards. (Presented by title.)

Abstracts of the contributed papers appear elsewhere in this issue of the Annals.

On Wednesday morning the Institute and the Econometric Society held a joint session on *Stochastic Processes* with Professor Harold Hotelling of the University of North Carolina presiding. Attendance was approximately ninety. Professor Hotelling presented an *Historical Summary of the Problem* Professor J. L Doob of the University of Illinois presented a paper, *Stochastic Differences Equations and Stochastic Differential Equations.* Professor Subrahmanyan Chandrasekhar of the University of Chicago presented a paper, *Brownian Motion, Dynamical Friction and Stellar Dynamics.*

The three joint sessions of the Institute and the Econometric Society on Thursday were devoted to a *Symposium on the Theory of Games*. The maximum attendance was approximately three hundred The first morning session was held under the chairmanship of Professor S S Wilks of Princeton University.

Professor John von Neumann of the Institute for Advanced Study presented a paper, *Survey of the Theory of Games.* Professor Oskar Morgenstern of Princeton University presented a paper, *Economics and the Theory of Games.* Dr. M. A. Girshick of Project Rand presented a paper, *Statistics and the Theory of Games.* The second morning session was under the chairmanship of Professor John von Neumann of the Institute for Advanced Study. Dr. E. W. Paxson of Project Rand presented a paper, *Recent Developments.* Professor J. W. Tukey of Princeton University presented a paper, *A Problem in Strategy.* Dr. G. B. Dantzig of the Army Air Forces presented a paper, *Programming in a Linear Structure.* The final session of the symposium was a round table discussion with Professor John von Neumann of the Institute for Advanced Study as chairman and with the following participants: Dr. G. B. Dantzig, Dr. M. A. Girshick, Professor Harold Hotelling, Professor Irving Kaplansky, Professor Samuel Karlin, Dr. J. C. C. McKinsey, Professor Oskar Morgenstern, Dr. E. W. Paxson, Dr. L. S. Shapley, and Professor J. W. Tukey.

A membership business meeting was held on Friday, September 10, in Bascom Hall at which twenty-one members were present. An account of the business transacted at this meeting may be found elsewhere in this issue under the heading "Adoption of a New Constitution."

The final session was on *Sequential Estimation* and was held jointly with the Econometric Society on Friday morning with Professor Jerzy Neymen of the University of California presiding. Attendance was approximately fifty. Professor Charles Stein of the University of California presented a paper on *Sequential Estimation.* Professor W. A. Wallis of the University of Chicago presented a discussion.

Social affairs during the meeting included a tea Tuesday afternoon, a concert of the Pro Arte String Quartet Tuesday evening, a dinner Wednesday evening, a picnic Thursday afternoon, and a beer party Thursday evening.

K. J. Arnold
Assistant Secretary

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 19[illegible])

BOARD OF DIRECTORS FOR 19[illegible]

President:

A. WALD, Columbia University, New York 27

Vice-Presidents:

CHURCHILL EISENHART, National Bureau of Standards, Washington [illegible]

HENRY SCHEFFÉ, Columbia University, New York 27

Secretary-Treasurer:

PAUL S. DWYER, University of Michigan, Ann Arbor

Past Presidents:

WILL FELLER, Cornell University, Ithaca, N. Y.

WILLIAM G. COCHRAN, North Carolina State College, Raleigh

Editor:

S. S. WILKS, Princeton University, Princeton, N. J.

The purpose of the Institute of Mathematical Statistics is to stimulate research in the mathematical theory of statistics and to promote cooperation between the field of pure research and the fields of application.

Membership dues including subscription to the ANNALS OF MATHEMATICAL STATISTICS are $7.00 per year within the Western Hemisphere and $[illegible] per year elsewhere. Dues and inquiries regarding membership in the Institute should be sent to the Secretary-Treasurer of the Institute.

MEETINGS OF THE INSTITUTE

ANNUAL MEETING—CLEVELAND, OHIO, December [illegible]

To be held in conjunction with meetings of the American Statistical Association, the Biometric Society, the Econometric Society and related social science organizations.

Abstracts must be in the hands of Secretary P. S. Dwyer, [illegible] Hall, University of Michigan, Ann Arbor, Michigan, not later than November 15.

NEW YORK CITY, April 8–9, 1949

Abstracts must be in the hands of Assistant Secretary [illegible], Department of Industrial Engineering, Columbia University, New York 27, N. Y., not later than March 15.

BERKELEY, June 17, 1949

Abstracts must be in the hands of Assistant Secretary, [illegible] Laboratory, University of California, Berkeley [illegible], not later than May 15.

BOULDER, COLO., August 29–September [illegible]

Abstracts must be in the hands of the Secretary not later than [illegible]

LOS ANGELES, November 25–26, 1949

Abstracts must be in the hands of the Secretary not later than [illegible]

ANNUAL MEETING—NEW YORK CITY, December [illegible]

To be held in conjunction with the meetings of the American Statistical Association.

Abstracts must be in the ha[illegible] of Assistant Secretary [illegible], Department of Industri[illegible], Columbia University, New York 27, N. Y., not later t[illegible]